**Replication-dependent biogenesis of turnip crinkle virus long noncoding RNAs**

Shaoyan Zhang, Rong Sun, Liming Zheng, and Feng Qu

Department of Plant Pathology, Ohio Agricultural Research and Development Center, The Ohio State University, 1680 Madison Ave, Wooster, OH 44691

Contact: Feng Qu, qu.28@osu.edu; 330-263-3835

**Abstract**

Long noncoding RNAs (lncRNAs) of virus origin accumulate in cells infected by many positive strand (+) RNA viruses to bolster viral infectivity. Their biogenesis mostly utilizes exoribonucleases of host cells that degrade viral genomic or subgenomic RNAs in the 5'-to-3' direction until being stalled by well-defined RNA structures. Here we report a viral lncRNA that is produced by a novel replication-dependent mechanism. This lncRNA corresponds to the last 283 nucleotides of the turnip crinkle virus (TCV) genome, hence is designated tiny TCV subgenomic RNA (ttsgR). ttsgR accumulated to high levels in TCV-infected *Nicotiana benthamiana* cells when the TCV-encoded RNA-dependent RNA polymerase (RdRp), also known as p88, was overexpressed. Both (+) and (-) strand forms of ttsgR were produced in these cells in a manner dependent on the RdRp functionality. Strikingly, templates as short as ttsgR itself were sufficient to program ttsgR amplification, as long as the TCV-encoded replication proteins, p28 and p88, were provided *in trans*. Consistent with its replicational origin, ttsgR accumulation required a 5' terminal $G_3(A/U)_4$ motif shown by others to be crucial for the replication of a TCV satellite RNA. More importantly, introducing a new $G_3(A/U)_4$ motif elsewhere in the TCV genome was alone sufficient to cause the emergence of another lncRNA.

Collectively our results unveil a replication-dependent mechanism for the biogenesis of viral lncRNAs, thus suggesting that multiple mechanisms, individually or in combination, may be responsible for viral lncRNA production.

**Importance**

Many positive strand (+) RNA viruses produce long noncoding RNAs (lncRNAs) during the process of cellular infections, and mobilize these lncRNAs to counteract antiviral defenses, as well as coordinate the translation of viral proteins. Most viral lncRNAs arise from 5'-to-3' degradation of longer viral RNAs being stalled at stable secondary structures. We report a viral lncRNA that is produced by the replication machinery of turnip crinkle virus (TCV). This lncRNA, designated ttsgR, shares the terminal characteristics with TCV genomic and subgenomic RNAs, and over-accumulates in the presence of moderately overexpressed TCV RNA-dependent RNA polymerase (RdRp). Furthermore, templates that are of similar sizes as ttsgR itself are readily replicated by TCV replication proteins (p28 and RdRp) provided from non-viral sources. In summary, this study establishes an approach for uncovering low abundance viral lncRNAs, and characterizes a replicating TCV lncRNA. Similar investigations on human-pathogenic (+) RNA viruses could yield novel therapeutic targets.

**Introduction**

Viruses with single-stranded, positive sense (+) RNA genomes are among the most common pathogens of humans, animals, and plants [1]. Well known examples of human-pathogenic (+) RNA viruses encompass poliovirus, Dengue virus, Zika virus, and most recently the SARS-CoV-2 responsible for the ongoing COVID-19 pandemic [2–5]. Plant-infecting (+) RNA viruses include tobacco mosaic virus (TMV), the first virus ever discovered, and many other agriculturally important crop pathogens. Despite their vastly different host tropism, (+)

RNA viruses share many important characteristics. For example, they all use their genomic RNAs as mRNA to direct the translation of at least some of the virus-encoded proteins, and replicate their genomes via negative strand (-) replication intermediates. As a result, research on both animal and plant-infecting (+) RNA viruses have historically contributed to our understanding of these shared characteristics, and informed on strategies for control and management of diseases caused by these viruses.

A recently re-discovered characteristic shared by animal and plant-infecting (+) RNA viruses is the accumulation of long noncoding RNAs (lncRNAs) of virus origin in cells they infect. LncRNAs are generally defined as RNAs that are more than 200 nucleotides (nt) long, yet lack recognizable coding capacities [6,7]. Viral lncRNAs have been observed in (+) RNA virus-infected plants for a long time [8–12]. Similarly, many animal viruses of the family *Flaviviridae* were also reported to accumulate lncRNAs in infected cells [13–15]. These viral lncRNAs are almost always co-terminal with the 3' ends of the genomic RNAs of cognate viruses. However, the questions of how these viral lncRNAs are produced, as well as what roles they play in viral infections, remained unresolved until 2008, when lncRNAs of animal-infecting flaviviruses, and that of plant-infecting red clover necrotic mosaic dianthovirus (RCNMV), were independently established as products of stalled 5'-to-3' degradation of longer viral RNAs by exoribonucleases encoded in host cell genomes, with the stalling caused by well-defined RNA secondary structures within these lncRNAs [16,17].

Extensive follow-up investigations have not only revealed the existence of similar viral lncRNAs in many more viruses, but also identified the primary 5'-to-3' exoribonucleases – Xrn1 in animals and XRN4 in plants – responsible for their biogenesis [18–27]. The specific RNA structures that act to stall Xrn1 or XRN4 degradation have been thoroughly examined, and in a

few cases the crystal structures resolved [19,28–32]. These lncRNA-resident structures have complex modular features including multiple stem-loops interspersed with frequent, and often long-distance pseudoknot interactions.

Perhaps the most astounding revelation from these investigations is that these viral lncRNAs, despite their replication-independent biogenesis, all have pro-viral functions during the life cycles of their parental viruses [33]. These functions ranged from mitigating host antiviral defenses such as RNA interference (RNAi), interferon signaling, and other forms of innate immunity; to regulating the translation of various virus-encoded proteins [17,19,23]. Preservation of these important functions by viral lncRNAs in diverse (+) RNA viruses raises an interesting question: could some viruses have evolved the ability to produce similar lncRNAs via *de novo* synthesis and/or amplification by virus-encoded replication proteins? Namely, could functional lncRNAs of some viruses be products of virus-mediated replication or transcription?

Here we report the identification and characterization of a lncRNA derived from turnip crinkle virus (TCV). TCV is a small plant-infecting (+) RNA virus with a genome of 4,054 nucleotides (nt), encoding five proteins [34]. The two proteins directly translated from the genomic RNA (gRNA) are the p28 auxiliary replication protein and the p88 RNA-dependent RNA polymerase (RdRp), with the latter being the product of infrequent translational readthrough overcoming the p28 stop codon (Fig. 1a). TCV also produces two subgenomic RNAs (sgRNAs) during cellular infections that encode the movement proteins p8 and p9, and coat protein (CP) p38, respectively. p38 is also the suppressor of RNA silencing encoded by TCV [35–37]. The existence of a TCV-encoded lncRNA, designated tiny TCV subgenomic RNA (ttsgR), was first unveiled by overexpressing p88 in cells replicating a TCV replicon [38]. In the current study, we further examined the biogenesis of ttsgR independent of a replicating TCV

replicon, and discovered a replication-based mechanism for ttsgR production. Our results revealed an alternative viral lncRNA biogenesis strategy that does not involve 5'-to-3' exonuclease degradation of longer viral RNAs.
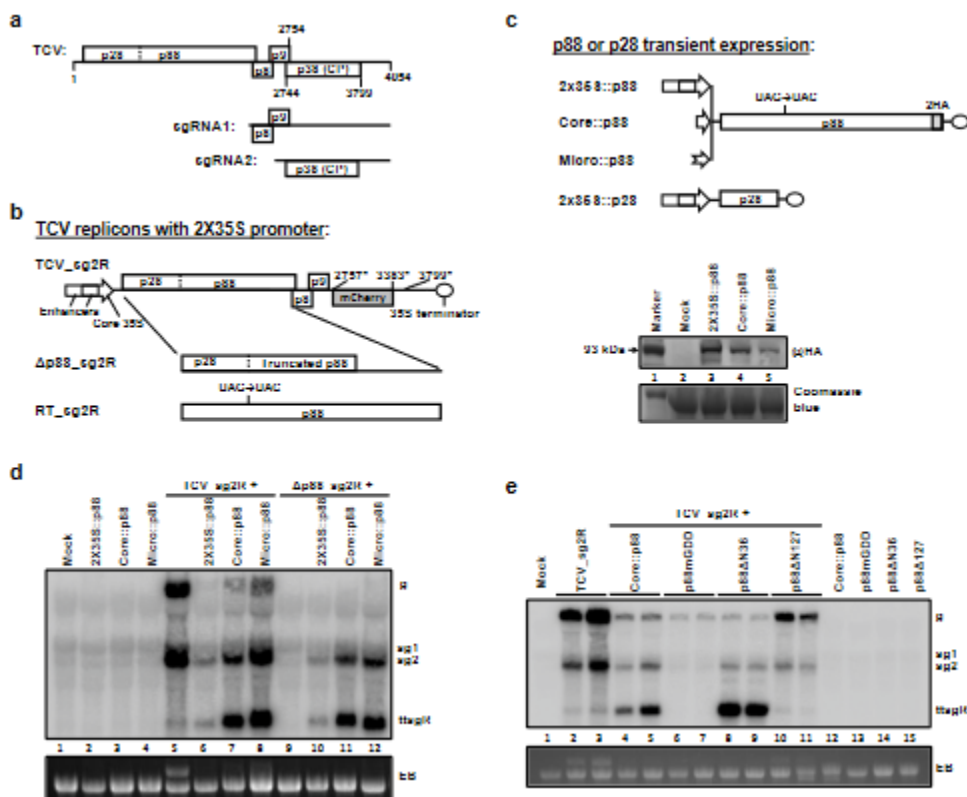


**Figure 1. Moderately overexpressed p88 stimulates abundant ttsgRNA accumulation.**

**a.** Schematic representation of the TCV genome with encoded proteins shown as rectangles. The two sgRNAs, sgRNA1 and 2, are produced during cellular infections, and serve as mRNA for p8 and p9, and p38 (CP), respectively. **b.** TCV replicons used in the current study. TCV_sg2R encodes wildtype (wt) p28 and p88, but expresses the red fluorescent mCherry instead of p38 from sgRNA2. Δp88_sg2R contains a 4-nt deletion in the middle of p88 ORF, making it incapable of translating a functional p88. Conversely, RT_sg2R contains a UAC codon in place of the p28 stop codon (UAG), rendering it incapable of translating p28, yet overexpressing p88. All three replicons use the strong 2X35S promoter of CaMV to drive the transcription of

infectious TCV RNA. **c.** Constructs that expresses TCV p88 and p28 independent of replication. Three promoters, 2X35S, Core35S, and Micro, were used to drive the expression of p88. Note that p88 expressed from these constructs contain a C-terminal, duplicated HA (2HA) tag that permits protein detection with an HA antibody. This tag does not affect the function of p88 [38]. By contrast, the p28-expressing construct contained a 2X35S promoter preceding the p28 ORF, but no epitope tag at the C-terminus. The image beneath the diagrams shows a Western blotting detection of the 2HA-tagged p88 from cells of *N. benthamiana* receiving the three p88 constructs. **d.** Repression and replicational complementation of TCV replicons by p88 are accompanied by preferential over-accumulation of ttsgR and sgRNA2. The Northern blotting autoradiograph shows the accumulation levels of TCV gRNA (g), sgRNAs (sg1 and 2), and ttsgR in the total RNA samples extracted from *N. benthamiana* leaves that were treated with various constructs or combinations thereof. Note that a TBSV p19-expressing construct was always included to mitigate RNA silencing-mediated mRNA degradation. EB: ethidium bromide stained agarose gel serving as the loading control. **e.** The RdRp activity of p88 is required for ttsgR over-accumulation.

## Results

**ttsgR accumulates to high levels in the presence of moderately overexpressed p88.** We recently reported that replication of a TCV replicon, TCV_sg2R, was inhibited by the overexpression of p88, the TCV-encoded RdRp [38]. TCV_sg2R encodes the red fluorescent mCherry protein in place of TCV p38 [38,39] (Fig. 1b), permitting convenient monitoring of its replication with confocal microscopy. Although TCV_sg2R was unable to produce the p38 RNA

silencing suppressor [35], it replicated robustly in *N. benthamiana* cells when supplemented with

p19 of tomato bushy stunt virus (TBSV). However, TCV_sg2R replication was repressed by p88

translated from a separate, non-replicating mRNA [38]; and this repression inversely correlated

with the p88 protein levels [38–40]. Intriguingly, ttsgR, a small TCV-derived RNA present at

low levels in regular TCV_sg2R infections (Fig. 1d, lane 5), became highly abundant in the

presence of moderate levels of p88 provided with the Core::p88 construct (Fig. 1c, bottom, lane

4; and Fig. 1d, lane 7). The Core promoter in this construct contained the last 99 nt of the

cauliflower mosaic virus (CaMV) 35S promoter, devoid of the upstream enhancer (Fig. 1b). It

drove a lower expression of p88 than the 2X35S::p88 construct (Fig 1c, bottom, lanes 3 and 4).

As a result, Core::p88 was less potent than 2X35S::p88 at repressing TCV_sg2R replication (Fig.

1d, lanes 6 and 7). Conversely, it was more competent at complementing the replication of the

p88-defective Δp88_sg2R replicon (Fig. 1d, lanes 10 and 11), indicating that lowering p88

expression made it more replication-competent [38]. Accordingly, the high abundance of ttsgR in

lanes 7 and 11 of Fig. 1d, in comparison with lanes 6 and 10, could be interpreted as preferential

replication/transcription of shorter TCV RNAs in the presence of moderately overexpressed p88.

This interpretation would be consistent with the concurrent overaccumulation of TCV sgRNA2

(Fig. 1d, lanes 6, 7 and 10, 11).

If this interpretation was correct, we would expect the level of ttsgR to increase even more

if p88 levels were further decreased. To test this prediction, we next expressed p88 using a

tobacco promoter known to be microspore-specific [41], hence only minimally active in *N.

benthamiana* leaf cells (Micro::p88. Fig. 1c, top). The Micro::p88 construct indeed expressed a

very low level of p88 protein (Fig. 1c, bottom, lane 5), though probably still higher than p88

produced by readthrough translation from replicon genomes. Accordingly, its repression of

TCV_sg2R replication was weaker than Core::p88 (Fig. 1d, lane 8). However, it was only marginally better at complementing the replication of Δp88_sg2R (Fig. 1d, lane 12). Nevertheless, in both cases the ttsgR levels, along with that of sgRNA2, visibly increased. These results were consistent with the involvement of the p88 RdRp activity in ttsgR biogenesis.

To directly assess whether the RdRp activity of p88 was needed for ttsgR accumulation, we created p88mGDD, an RdRp-null mutant of p88, by changing the highly conserved GDD motif to VAA in the Core::p88 background. As shown in Fig. 1e (lanes 6 and 7), expression of p88mGDD caused a dramatic reduction of TCV_sg2R gRNA and sgRNAs, and disappearance of ttsgR, recapitulating only the repressive activity of overexpressed p88 [38]. This result contrasted with the selective repression of gRNA by Core::p88 that was accompanied by preferential accumulation of ttsgR and sgRNA2 (lanes 4 and 5), hence implicating the RdRp activity of p88 in ttsgR overaccumulation.

To further interrogate the role of p88 RdRp in ttsgR biogenesis, we additionally tested two deletion mutants of Core::p88, p88ΔN36 and p88ΔN127, that removed the N-terminal 36 and 127 amino acid (aa) residues, respectively, from the p88 open reading frame (ORF) [38]. We previously established that while p88ΔN36 retained both the repressive and complementing activities of p88 to a substantial extent, p88ΔN127 was modestly repressive but incapable of complementing p88-defective replicons. As shown in Fig. 1e, preferential over-accumulation of ttsgR occurred in the presence of p88ΔN36 (lanes 8 and 9), but not p88ΔN127 (lanes 10 and 11). Together these results provided further support for the involvement of p88 RdRp activity in ttsgR biogenesis.

Additional observations summarized in the Supplemental Fig. 1 (S1 Fig) corroborated the importance of p88 in ttsgR production. First, although p88 encompasses p28 at its N-terminus,

overexpression of the entire p88, rather than p28, was necessary for ttsgR overaccumulation (S1 Fig, lanes 2 – 5). Second, for ttsgR overaccumulation to occur, p88 overexpression did not have to be from a non-replicon source. Instead, abundant ttsgR (and sgRNA2) accumulated when the RT_sg2R mutant replicon was made to replicate with p28 provided *in trans* (S1 Fig, a, lanes 6-9). The RT_sg2R mutant lacked the p28 stop codon, causing it to sacrifice p28 production in favor of p88 overproduction (Fig. 1b). As a result, its replication could only occur in the presence of p28 provided *in trans*. Third, ttsgRNA overaccumulation was triggered by p88 overexpression during the replication of not only TCV_sg2R, but also wildtype (wt) TCV (S1 Fig, b, lanes 4-7). Interestingly, another TCV lncRNA slightly larger than ttsgR also accumulated in wt TCV infections (arrowhead in S1 Fig, b), the significance of which will be discussed.

**ttsgR is unlikely to be a degradation product of longer TCV RNAs.** Results in the previous section strongly hinted that ttsgR overaccumulation was caused by the RdRp activity of overexpressed p88. If true, this conclusion would be in stark contrast with lncRNAs associated with RCNMV and tobacco necrosis virus-D (TNV-D), which were both products of exonuclease-mediated 5'-to-3' degradation of larger viral RNAs [17,19]. This contrast is noteworthy because TCV is highly similar to both RCNMV and TNV-D in terms of genome organization and sequence homology of replication proteins, with all three of them belonging to the family *Tombusviridae*. It in turn demanded more rigorous scrutiny of our observations.

The lncRNA of TNV-D, known as svRNA, primarily arose from the degradation of TNV-D sgRNA1, one of its two sgRNAs[19]. Incidentally, ttsgR overaccumulation was frequently accompanied by preferential overproduction of TCV sgRNA2. To determine whether ttsgR abundance depended on either of the two TCV sgRNAs, we tested three TCV_sg2R mutants,

9

Δsg1, Δsg2, and Δsg1&2, in which the production of sgRNA1, sgRNA2, or both, were abolished through mutagenesis of sgRNA promoters [42,43] (Fig. 2a). As shown in Fig. 2b, none of the three mutants abolished p88-dependent ttsgR accumulation, despite the loss of either or both sgRNAs in the corresponding mutants. Therefore, ttsgR was unlikely the degradation product of TCV sgRNAs.

We next examined whether ttsgR could have been derived from degradation of TCV gRNA. For this purpose we generated three new constructs: [p28stop]_sg2R, sg2R_Temp, and ttsgR_Temp (Fig. 2c). Note that while [p28stop]_sg2R was a full-length replicon, the extra stop codon at aa position 36 of p28 prevented the production of both p28 and p88, making it incapable of launching replication by itself. The sg2R_Temp and ttsgR_Temp constructs were likewise incapable of self-replication because their transcripts, 1,671 and 669 nt respectively, were 3' co-terminal fragments of TCV_sg2R gRNA with their 5' ends located at nt #2,468 and 3,386, both downstream of the p28 and p88 ORFs (Fig. 2c). However, when p28 and p88 were provided *in trans* (with constructs 2X35S::p28 and Core::p88; Fig. 1c), all three constructs produced TCV-specific RNAs of expected sizes (Fig. 2d and S2 Fig). Importantly, accumulation of these RNAs depended on the presence of both p28 and p88, and required the GDD motif of p88 RdRp (Fig. 2d and S2 Fig). Most significantly, all three templates additionally accumulated ttsgR in a manner dependent on the presence of both p28 and p88. Therefore, ttsgR was independent of the degradation of full-length TCV gRNA. On the contrary, ttsgR biogenesis strictly required the RdRp activity of p88 (Fig. 2d; S2 Fig), and could occur with a template as short as the last 669 nt of TCV gRNA.
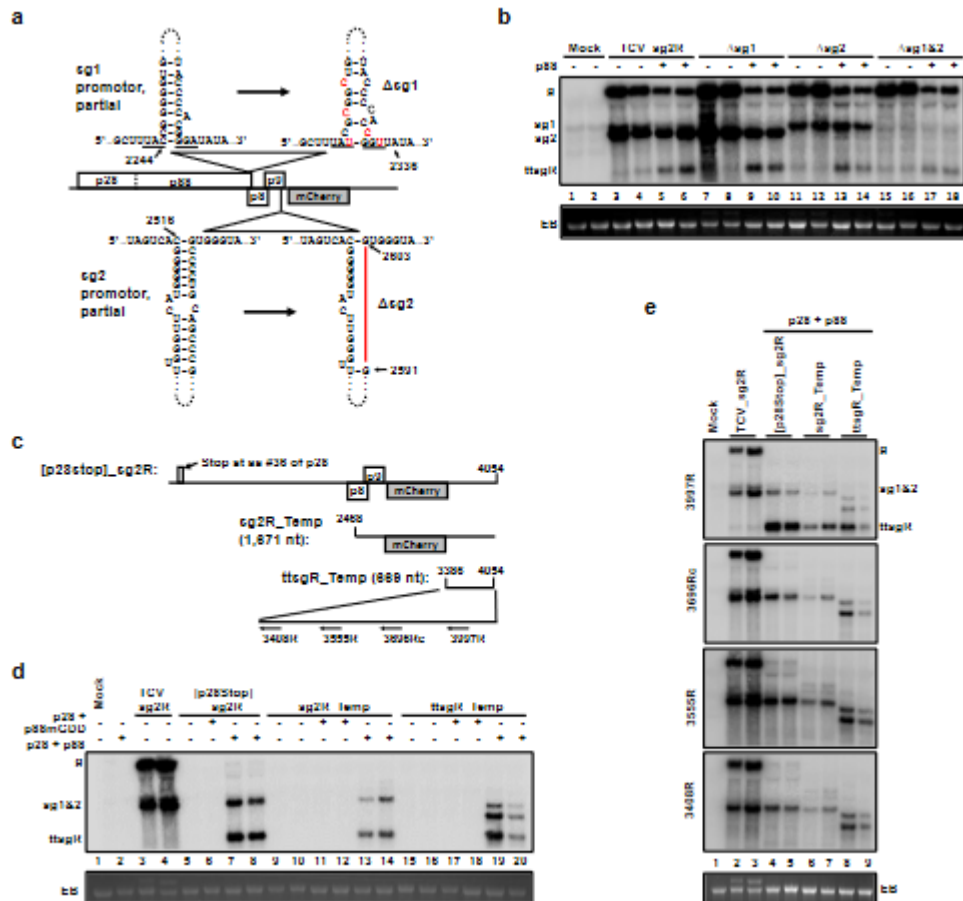
**Figure 2. ttsgR is unlikely a product of 5'-to-3' degradation of longer TCV RNAs. a.**
Schematic depiction of Δsg1 and Δsg2 mutants of TCV_sg2R. The bottom portion of the stem-
loop structure identified previously as the sgRNA1 promoter is shown on the top left. The Δsg1
mutant, with the five single nt mutations shown in red letters, is represented to the right.
Similarly the partial sgRNA2 promoter stem-loop is at the bottom left, with the corresponding
Δsg2 mutant shown to the right, in which a deletion of 12 nt is denoted as a red line. **b**. ttsgR
accumulation was not abolished in Δsg1, Δsg2, and Δsg1&2 mutants, even though the levels of
the corresponding sgRNAs were drastically diminished, as evidenced by Northern blotting
results. **c**. Schematic representation of three template constructs that transcribe 3'-coterminal
TCV (TCV_sg2R)-derived RNAs of varying lengths, with none of them capable of translating

the replication proteins. **d**. The three defective templates templated the production of ttsgRNA, in addition to RNAs of template lengths, in the presence of trans-supplied p28 and p88, but not p28 and p88mGDD. Note that the [p28stop]_sg2R template accumulated very low levels of full length RNA, but a easily detectable level of sgRNA2 and ttsgR. This outcome was expected as overexpressed p88 complemented the replication of TCV gRNA poorly. **e**. Preliminary mapping of the 5' terminus of ttsgR by Northern blotting with four different probes.

We next tried to pinpoint the 5' terminus of ttsgR by subjecting these RNA samples to Northern blot hybridizations with four different oligonucleotide probes (Fig. 2c, bottom). As shown in Fig. 2e, ttsgR was readily detectable with the 3997R oligo probe, but not with the 3696Rc probe, or two other probes that hybridized to more upstream positions of TCV gRNA. These experiments narrowed down the 5' end of ttsgR to a position approximately between positions 3,690 and 3,997 of TCV, suggesting a ttsgR size of no more than 365 nt.

**ttsgR initiates at a discrete position of TCV RNA, and it exists in both (+) and (-) strand forms.** To ultimately prove ttsgR as the product of p88 RdRp activity, we needed to map its terminal sequences, and ideally demonstrate that a template with the exact termini of ttsgR amplifies itself in the presence of p88 (and p28). To these ends, we first mapped the termini of ttsgR using a circularization RACE procedure [44]. Briefly, this procedure used T4 RNA ligase to connect the 5' end of an RNA to its own 3' end, hence circularize the RNA (Fig. 3a). The circularized RNA was then subjected to reverse transcription (RT)-polymerase chain reaction (PCR) with divergent primers, followed by sequencing to discern the sequence of the 3'-5' junction (Fig. 3a). Hence, this approach would allow us to map the 5' and 3' ends of ttsgR simultaneously. To further determine whether both (+) and (-) strands of ttsgR were present in

the samples, we adopted a strand-specific (ss) RT-PCR procedure that utilized non-viral

sequence tags (Tag and Tg2 in Fig. 3a and b; S4 Table) to minimize the contamination of
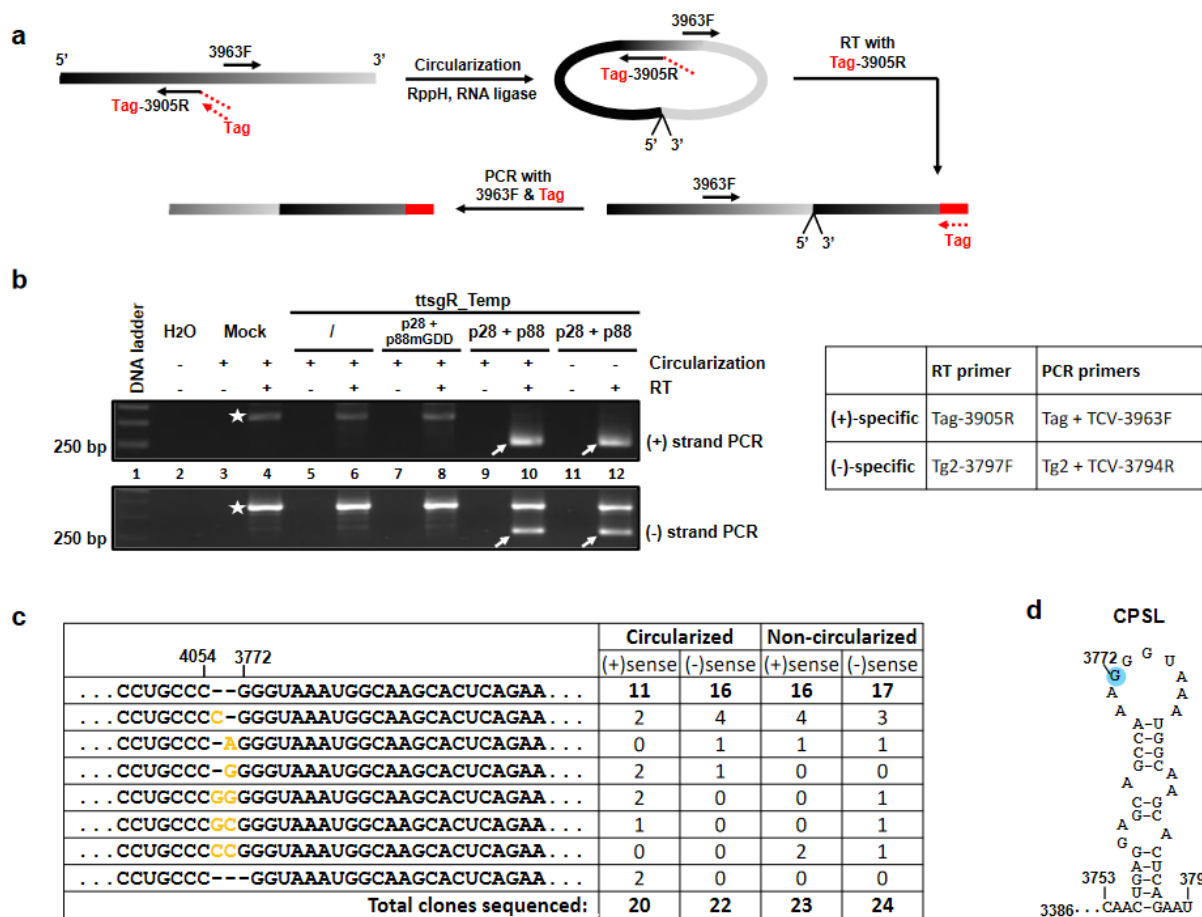
opposite strands [45].



**Figure 3**. The 5' terminus of ttsgR maps to nt #3,772 of TCV gRNA. **a**. Step-by-step

depiction of the Circularization RACE. Note that (i) a non-TCV tag was fused to the 5' end of

the RT primer to enhance the strand-specificity; and (ii) the two PCR primers (Tag and 3963F)

are back-to-back from each other, hence a PCR product is only possible when the template is

circularized or in dimeric form. **b**. PCR products of both (+) and (-) strands arisen from

Circularization RACE. The stars on both agarose gels denote non-specific products present even

in mock-treated samples. The primers used are shown to the right. **c**. Sequencing results of

cloned PCR products (arrows in **b**) delimit the starting and ending nt of ttsgR. **d**. The position of

the ttsgR 5' terminus in a previously identified origin-of-assembly stem-loop within the CP ORF

(CPSL).

As shown in Fig. 3b, specific, RT-dependent fragments were detected in samples in which

ttsgR_Temp was administered along with p28 and p88, but not p28 and p88mGDD (compare

lanes 8, 10, and 12). Furthermore, both (+) and (-) strands of ttsgR were successfully amplified.

Surprisingly, RNA circularization was unnecessary for the amplification of both (+) and (-)

strands, indicating that at least some of the ttsgR molecules existed as tandem dimers. The

presence of (-) ttsgR, coupled with detection of dimers, is consistent with ttsgR undergoing

active replication in the presence of p28 and p88.

The PCR products highlighted by arrows (Fig. 3b) were purified and cloned. For each of

the PCR products, more than 20 individual clones were sequenced. As shown in Fig. 3c,

regardless of the sense of the RNA, or whether the RNA was circularized, more than 50% of

clones in each category shared a 3'-5' junction that connected the 3' end CCC with GGG

beginning with nt #3,772 of TCV. The rest of clones mostly contained 1-2 extra G or C between

the two termini. It should be noted that because circularized and non-circularized RNA samples

yielded similar sequencing results, most of the junction sequences were probably derived from

dimeric RNAs. As a result, some of the RNA species, especially those with rare junction

identities, might have replicated as dimeric rather than monomeric form. Nevertheless, the fact

that most clones had #3,772 joining #4,054 led us to conclude that the ttsgR is predominantly

283 nt long, most likely arose from RdRp-mediated amplification that initiated at nt #3,772.

Notably, this initiation site is located within the end loop of a stem-loop structure previously

identified as important for the packaging of TCV gRNA in virions [46] (Fig. 3d), prompting

further investigations to determine whether the same RNA structure was needed for ttsgR

biogenesis.

**The hairpin structure flanking the 5' end of ttsgR is not needed for ttsgR biogenesis.**

That ttsgR initiated at a position within the end loop of a previously identified hairpin structure

[46] prompted us to evaluate the importance of this structure in ttsgR biogenesis. This was first

carried out in the ttsgR_Temp background, which allowed us to disrupt the structure extensively.

As shown in Fig. 4a, the ttsgR_mL and -mR mutants each contained eight single-nt mutations (in

red letters) that disrupted nearly all base pairs in the stem, whereas the ttsgR_mLR combined the

mutations in the first two mutants to restore the base pairing in the stem, but not the original

sequence. However, none of the three mutants abolished ttsgR accumulation. Therefore, this

stem-loop structure was not required for ttsgR production from the ttsgR_Temp template.

We next tried to destabilize this structure, designated CP-resident stem-loop or CPSL, in

the contexts of the TCV_sg2R replicon and wildtype (wt) TCV. The new mCPSL mutants had to

preserve the identity of aa residues of TCV CP (p38), hence contained just six single nt changes

(Fig. 4c). Nevertheless, the CPSL was predicted to be substantially weakened. However, these

changes failed to compromise p88-dependent ttsgR accumulation in either TCV_sg2R or wt

TCV infections. Therefore, the stem-loop structure surrounding the ttsgR initiation site is not
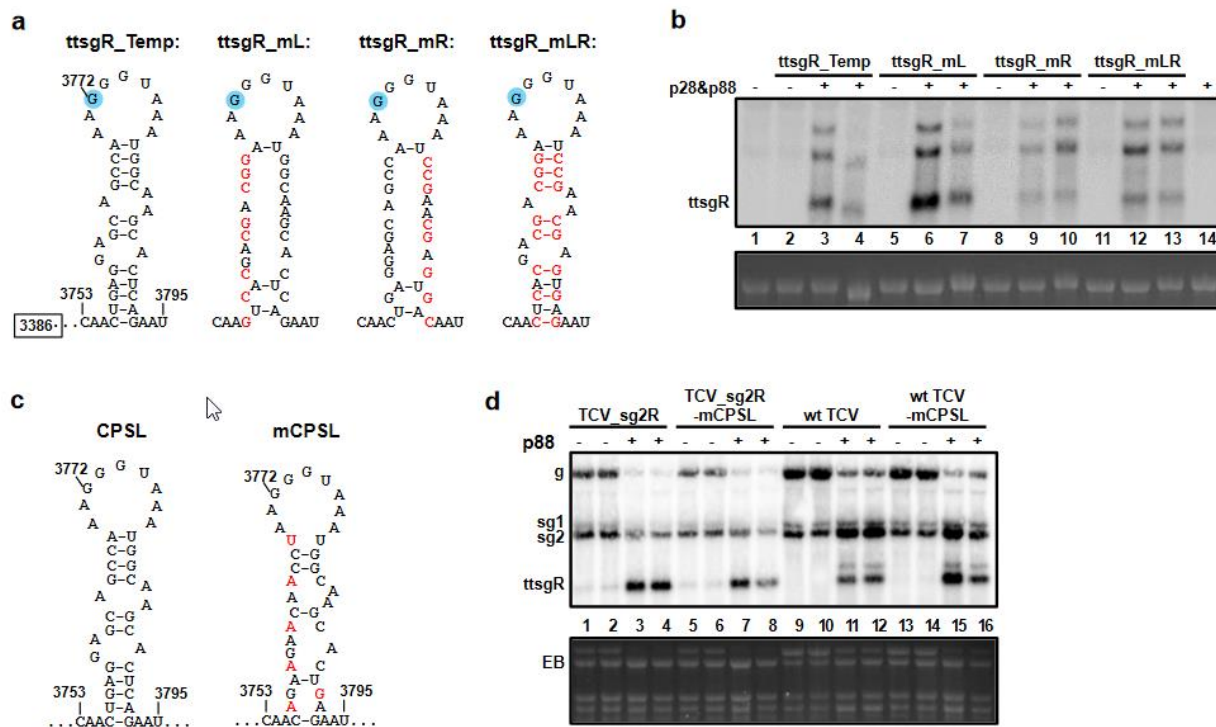
needed for ttsgR biogenesis from replicating TCV either.

**Figure 4**. Disruption of the CPSL structure has no effect on ttsgR production from either the ttsgR_Temp (Fig. 2c), or TCV replicons. **a**. Specific mutations (red letters) that disrupt CPSL in the ttsgR_Temp construct. **b**. p28&p88-dependent ttsgR production from various ttsgR_Temp mutants detected with Northern blotting. **c**. Mutations that disrupt CPSL in TCV replicons. **d**. p88-dependent production of ttsgR from the mCPSL mutants of TCV_sg2R and wt TCV.

**A template with 28 extra 5' end nucleotides reveals a specific terminal motif for ttsgR.**

The precise mapping of the ttsgR 5' end allowed us to test a shorter template, referred to as ttsgR_S, that begins with TCV nt #3,744, with just 28 extra nt upstream of the ttsgR initiation site at nt #3,772 (Fig. 5a). As shown in Fig. 5b (lanes 6-7), ttsgR_S permitted p28/p88-dependent accumulation of RNAs similar to ttsgR in size (compare lanes 6-7 with 3-4). Curiously, a slightly larger RNA species, designated ttsgR*, was also detected (arrows in lanes 6 and 7). In addition, another substantially larger, but fainter species was detected as well (diamonds in lanes 6, 7, 9,

10), probably representing ttsgR/ttsgR* dimers. The termini of ttsgR and ttsgR* were mapped

with divergent pairs of primers designed to amplify their dimeric forms, hence bypassing RNA

circularization. Sequence analysis of the RT-PCR products verified that ttsgR* extended ttsgR

by 32 nt at the 5' end, initiating at a site within the vector plasmid that was 3-nt upstream of the

ttsgR_S insert (Fig. 5c). ttsgR and ttsgR* each accounted for four of the 18 clones sequenced

(Fig. 5c). The 5' ends of the remaining ten sequences congregated near that of ttsgR (three

clones), or ttsgR* (seven clones). These results indicated that replication proteins p28 and p88

successfully amplified two RNAs with two distinct initiation sites that were 32 nt apart.

Interestingly, the first seven nts of ttsgR and ttsgR*, GGGUAAA and GAGAAAU, appeared to

be distinctively similar: they began with GGG or GAG, followed by four A or U residues (Fig.
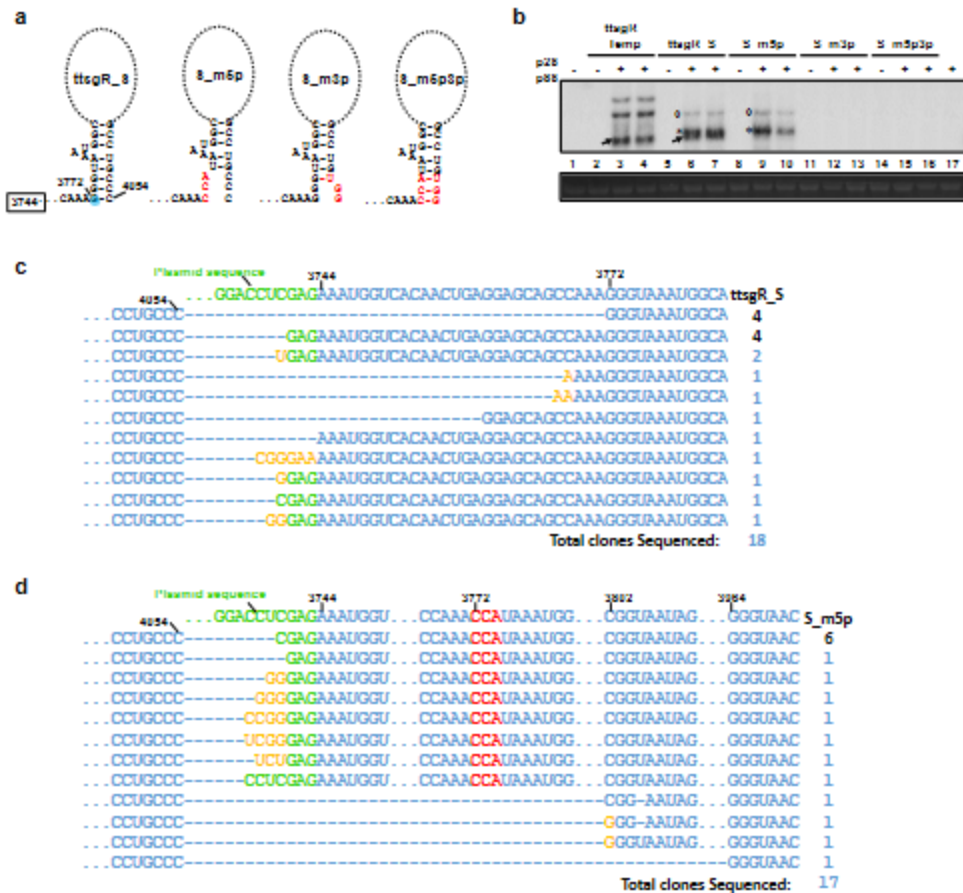
5c).

**Figure 5**. ttsgR_S, a template extending ttsgR by 28 nt at the 5' end, permits efficient ttsgR replication by p28 and p88 provided in *trans*, and facilitates the interrogation of potential 5'-3' base-pair interaction. **a**. Schematic representation of the ttsgR_S template, and three mutants of it, with mutated nts in red. The putative terminal stems formed by eight predicted base pairs, and its disruption by mutations, are also depicted. **b**. Northern blotting detection of ttsgR from five different templates. Arrows: ttsgR; stars: ttsgR*; diamonds: ttsgR/ttsgR* dimers. **c**. The terminal nt positions and identities of ttsgR variants arisen from the ttsgR_S template. **d.** The terminal nt positions and identities of ttsgR variants arisen from the S_m5p template. Data in **c** and **d** were obtained by sequencing the cloned products of RT-PCR primed with divergent (back-to-back) primers.

18

Closer inspection also revealed extensive sequence complementarity between 5' and 3' termini of ttsgR that could favor a double-stranded stem of eight GC-rich base pairs, with a 3-nt bulge (AAU) (Fig. 5a). We hence investigated whether this potentially stabilizing stem was critical for ttsgR accumulation, by first mutating the three basal base pairs (Fig 5a). All three mutants were based on ttsgR_S, hence designated S_m5p, S_m3p, and S_m5p3p (Fig. 5a). Interestingly, changing GGG at the 5' end of ttsgR to CCA (S_m5p) abolished ttsgR while appearing to leave the longer ttsgR* unchanged (Fig. 5b, lane 9 and 10). Sequence analysis verified that most of the clones harbored initiation sites at or near the upstream GAG, while none of them initiated at the original ttsgR initiation site (Fig. 5d). Therefore, the GGG or GAG motif at the 5' end appeared to be important for ttsgR and ttsgR* accumulation.

Intriguingly, three additional clones derived from the S_m5p mutant had their initiation sites at or near nt #3,802, 30 nt downstream of the authentic ttsgR initiation site (Fig. 5d, near bottom). Furthermore, the original sequence at this location, CGGUAAUA, became CGGAAUA, GGGAAUA, or GGGUAAUA, in the three clones. Therefore, while a sequence signature with two Gs followed by four U or As could also be used as an initiation site, the resulting replicating lncRNAs appeared to favor a GGG 5' end. Together, three classes of ttsgR_S and S_m5p replication products were represented by multiple clones, and all of them possessed characteristic 5' ends that could be summarized as $G_3(A/U)_4$, though the $G_3$ portion could tolerate some variations (e.g. GAG or GG).

Perhaps unsurprisingly, changing CCC to UGG at the 3' terminus (mutants S_m3p and S_m5p3p) completely abolished the accumulation of both RNA species (Fig. 5b, lanes 11-17). This result indicated that the identity of the CCC triplet was essential for the accumulation of

both ttsgR and ttsgR*. The 3' CCC triplet was previously found to be essential for the replication of TCV gRNA as well as a TCV-associated satellite RNA [47]. Therefore, the fact that the same CCC motif was essential for ttsgR accumulation provided further corroboration for a replicational origin of ttsgR.

**A ttsgR template with an authentic 5' terminus directs efficient ttsgR synthesis.** To further assess the involvement of any 5' end RNA structure, and stalled exonuclease degradation, in ttsgR accumulation, we then produced a third ttsgR template, ttsgR_S2, that has the same 5' terminus as ttsgR itself (Fig. 6a, left). Note that imprecise transcriptional initiation from the CaMV 35S promoter could generate template transcripts that contain up to 10 vector-derived nt (green letters in Fig. 6a). As shown in Fig. 6b (lanes 4-6), ttsgR_S2 permitted p28/p88-dependent accumulation of ttsgR. We further verified the identity of the accumulated lncRNAs by analyzing the sequences of RT-PCR products and found the expected ttsgR termini in seven of the 17 clones. The remaining ten clones, each representing one sequence, all contained slight variations at the 3'-5' junction. Note that some of them might represents variants that replicated exclusively in dimeric forms. Together these results demonstrated that a ttsgR template without any extra upstream TCV sequence or structure readily directed its own replication, and the replicated RNA is sufficiently stable to allow for consistent detection.
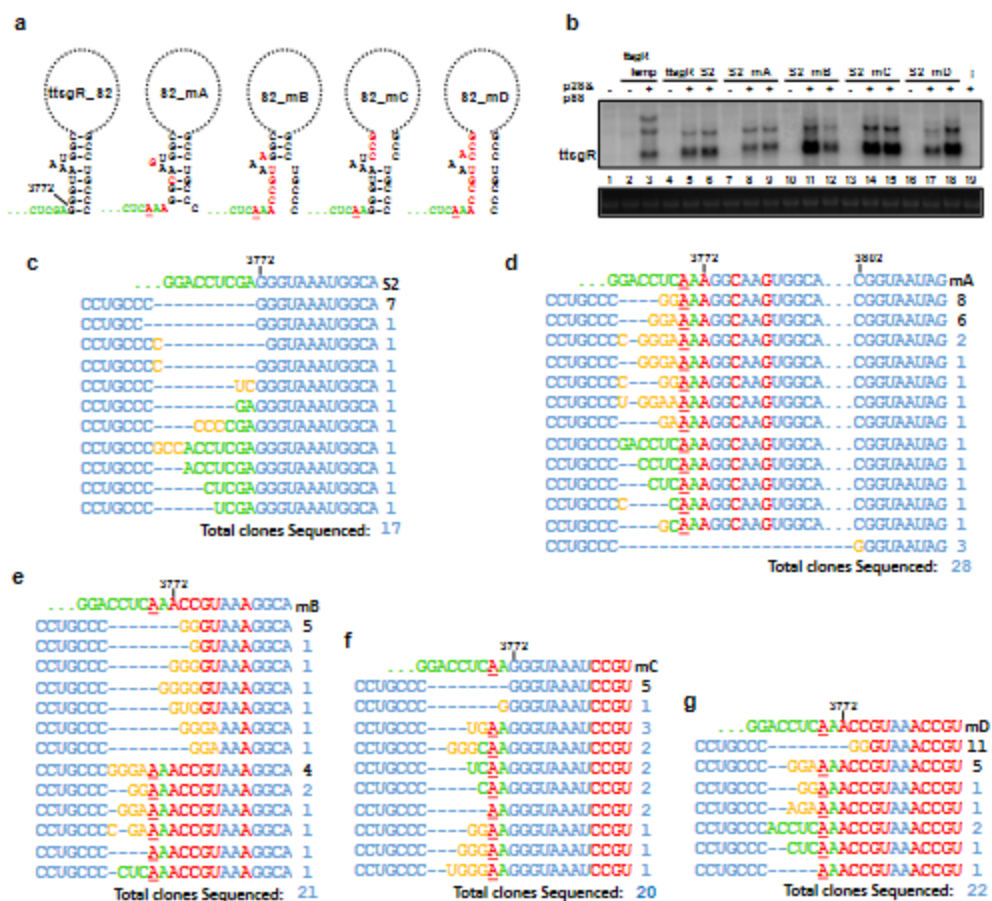
**Figure 6**. Interrogation of ttsgR_S2, a template with the exact ttsgR 5' terminus, reveals $G_3(A/U)_4$ as a conserved 5' end motif needed for replication of ttsgR variants. **a**. Schematic representation of the ttsgR_S2 template, and four mutants of it, with mutated nts in red. The putative terminal stems formed by eight predicted base pairs, and its disruption by mutations, are also depicted. **b**. Northern blotting detection of ttsgR from different templates. **c – g.** Terminal nt positions and identities of ttsgR variants arisen from the ttsgR_S2, and S2_mA – S2_mD templates. The sequence data were obtained by sequencing the cloned products of RT-PCR primed with divergent (back-to-back) primers.

**A 5' terminal G$_3$(A/U)$_4$ motif is critically important for ttsgR accumulation**. Since the

ttsgR_S2 template drove faithful amplification of ttsgR, it was next adopted as a template for

more extensive interrogations of potential terminal motifs that route template RNAs to

replication complexes containing p28 and p88. The ttsgR_S2 template also had the advantage of

minimizing initiation at the alternate GAG triplet (Fig. 6c), permitting more definitive

delineation of ttsgR 5' end. To determine whether the G$_3$(A/U)$_4$ motif at the ttsgR 5' terminus

tolerated certain variations, we generated four mutants in the sgR_S2 background, named as

S2_mA, S2_mB, S2_mC, and S2_mD, respectively (Fig. 6a). S2_mA contained three single nt

mutations in the G3(A/U)4 motif, changing it from GGGUAAU to AGGCAAG. It also

contained a G-to-A mutation in the nearby vector sequence, making it less likely to serve as an

alternate initiation site (Fig. 6a). Surprisingly, this mutant templated the synthesis of ttsgR

variants that migrated very slightly slower than ttsgR. Among the 28 cloned sequences analyzed,

most (25) retained the four mutations introduced (red-colored letters), but none of them initiated

at the mutated ttsgR initiation site (Fig. 6d). Instead, a majority of them acquired novel initiation

sites by incorporating non-template, non-vector nts that ranged from one to three Gs, and up to

two As (Fig. 6d, orange-colored letters). Together with the existing sequence (AAA) at the

vector-insert junction, these extra nt rebuilt the 5' ends that varied from GGAAA (8 clones),

GGAAAA (6 clones), to GGGAAAA (2 clones). Therefore, the mutated site had to be altered *de*

*novo* in order to serve as the initiation site for ttsgR replication, with the resulting sites

comprising at least two Gs, followed by at least three A/Us. Importantly, ttsgR variants with the

initiation site at nt #3,802 (following the numbering of full-length TCV genome) were again

recovered in three independent clones, with the first nt changed to G in all three clones (Fig. 6d,

bottom). Combined with similar results in Fig. 5d, these findings provided additional evidence for the importance of the $G_3(A/U)_4$ motif.

This necessity of the $G_3(A/U)_4$ motif was further reinforced by the next three mutants. The S2_mB mutant was originally designed to cause more extensive disruption to the putative long-distance terminal interaction shown in Fig. 5. It nevertheless also changed the GGGUAAA motif into ACCGUAA (Fig. 5A, middle diagram). However, there were still four U/As following the sole G residue. This mutant replicated robustly in *N. benthamiana* cells in the presence of p28 and p88 (Fig. 6b, lanes 10-12). Among the 21 cDNA sequences analyzed, approximately one half (11 clones; Fig. 6e, rows 2-8) initiated at the single G at position #3,775, but most of them also acquired additional Gs, or underwent other small modifications (e.g. U to A at #3,776, rows 7-8), to restore the $G_3(U/A)_4$ motif. The other half of the sequences (10 clones), like those of S2_mA, adopted the upstream AAA as the initiation site, but again acquired a varying number of Gs or Gs plus A to reconstitute the $G_3(U/A)_4$ motif. Notably, in these 10 sequences all of the original mutations were retained, suggesting that substantial 5'-3' base pairing, with the possible exception of three terminal G:C pairs, was not essential for ttsgR replication.

Many of the descendants of the mutant S2_mC appeared to deviate from the above generalization. Note that this mutant did not change the original GGGUAAA initiation motif, and it supported the accumulation of high levels of ttsgR (Fig. 6b, lanes 13-15). All of the 20 clones sequenced retained the original mutations downstream of GGGUAAA, indicating that the identity of these nts (and their potential base-pairing with the 3' end as depicted in Fig. 6a) were nonessential for ttsgR synthesis. RT-PCR products with the original ttsgR 5' terminus accounted for five of the 20 clones (Fig. 6f). The remaining 15 cDNAs contained insertions of up to six nt between the 3' CCC and 5' GGG, many of them being non-template nts. However, since the

S2_mC mutant was a highly potent ttsgR template, producing substantially higher levels of ttsgR than even the ttsgR_S2 template, it is possible that those with wildtype 5' ends accumulated predominantly in the monomeric form, hence were underrepresented in the pool of dimers we profiled. Conversely, those with insertions between CCC and GGG might have replicated in the dimeric form only.

Finally, the mutant S2_mD, which combined all mutations in S2_mB and S2_mC, also accumulated abundant ttsgRNA (Fig. 6b, lanes 16-18). Among the 22 clones sequenced, 11 contained cDNA sequences that initiated at the single G at nt #3,775 that preceded UAAA, but with two Gs added *de novo* (Fig. 6g). The other 11 sequences initiated from various upstream sites, with or without extra G/A added *de novo*. However, all of them contained at least three As in the vicinity of starting sites. Overall the data obtained with this set of mutants provided strong evidence for a conserved 5' end motif of $G_3(U/A)_4$. Intriguingly, the $G_3(U/A)_4$ motif at the 5' termini of the ttsgR variants highly resembled the consensus 5' termini shared by most carmovirus RNAs, including several TCV-associated satellite RNAs [48,49]. However, whether this motif alone could initiate the transcription/replication of TCV lncRNAs from full-length TCV replicons was not previously examined.

**The $G_3(U/A)_4$ motif is necessary and sufficient for initiating lncRNA synthesis from internal positions of TCV gRNA**. Results in previous sections demonstrated that a $G_3(U/A)_4$ motif is needed for p28/p88 to initiate the replication of ttsgR (and its variants) from templates that were extended at the 5' end by either 386 nt (ttsgR_Temp), 28 nt (ttsgR_S), or zero nt (ttsgR_S2) of TCV-derived sequences. Nevertheless, when the $G_3(U/A)_4$ motif in these templates was abolished through mutagenesis, similar motifs were relatively easily reconstituted in plant cells with spontaneous mutations or addition of non-template nt. However, it was not known

24

whether similar reconstitution could occur in full-length TCV replicons, e.g. TCV_sg2R or wt TCV. To resolve this question, we next tried to mutate the ttsgR-initiating $G_3(U/A)_4$ motif in both TCV_sg2R and wt TCV while maintaining the aa identity of CP. The first mutant, m1, altered three nts within the motif, changing G̲GGU̲AAA̲ to A̲GGC̲AAG̲ (Fig. 7a). In both TCV_sg2R and wt TCV, these mutations caused ttsgR to disappear even in the presence of overexpressed p88 (Fig. 7b, compare lanes 7, 8 with 3, 4; and 15, 16 with 11, 12). Although a new, low abundance lncRNA was produced by the TCV_sg2R-m1 mutant (Fig. 7b, lanes 7 and 8), it was slightly larger than ttsgR, probably initiating at the nt #3,740, from the GUGAAAA motif (see later). On the other hand, wt TCV produced another slower migrating lncRNA whose level was not affected by the m1 mutations (Fig. 7b, arrowheads). This lncRNA could have originated from GGGUUUA at nt #3,290, within the TCV CP coding region replaced by that of mCherry in TCV_sg2R (see discussion).

Next we mutated each of the three m1 mutations separately, creating mutants m1a, m1b, and m1c (Fig. 7a). As shown in Fig. 7c, while mutating the first (m1a; lane 8) or the last (m1c; lane 12) nt of the seven-nt motif had negligible effect on ttsgR accumulation, the U-to-C change immediately next to GGG (m1b) alone accounted for most of the loss caused by m1 in both TCV_sg2R and wt TCV (Fig. 7c and d, lanes 6 and 10). Therefore, the nt immediately following GGG (or GG) must be A or U in order to permit ttsgR initiation. Interestingly, while m1c itself had little effect on ttsgR abundance, another mutant obtained accidentally while producing m1c, referred as m1c-ii, caused a near complete loss of ttsgR in TCV_sg2R (Fig. 7c, lane 16). Since m1c-ii differed from m1c by just one nt – a deletion of A (Fig. 7a), this result suggested that a minimum of three U/A must follow GGG (or GG) in order to facilitate efficient ttsgR biogenesis. These results were also consistent with the sequencing results described in Figs. 5 and 6.

We then asked whether creation of a $G_3(A/U)_4$ motif at a different location of the TCV (and TCV_sg2R) genome was sufficient to cause the production of a new lncRNA. We hence mutated four nts between #3,401 and 3,407 of TCV, leading to a new $G_3(A/U)_4$ motif with the sequence of GGGUUAA (Fig. 7a). This set of m2 mutations were introduced into both TCV_sg2R and wt TCV, alone or with m1 mutations (Fig. 7a). As shown in Fig. 7e, the m2 mutations indeed led to the appearance a new lncRNA, but also simultaneously depressed the accumulation of ttsgR (lanes 6 and 7; 14 and 15). As expected, combining the m1 and m2 mutations led to a more complete loss of ttsgR, with little impact on the new, m2-dependent lncRNA.
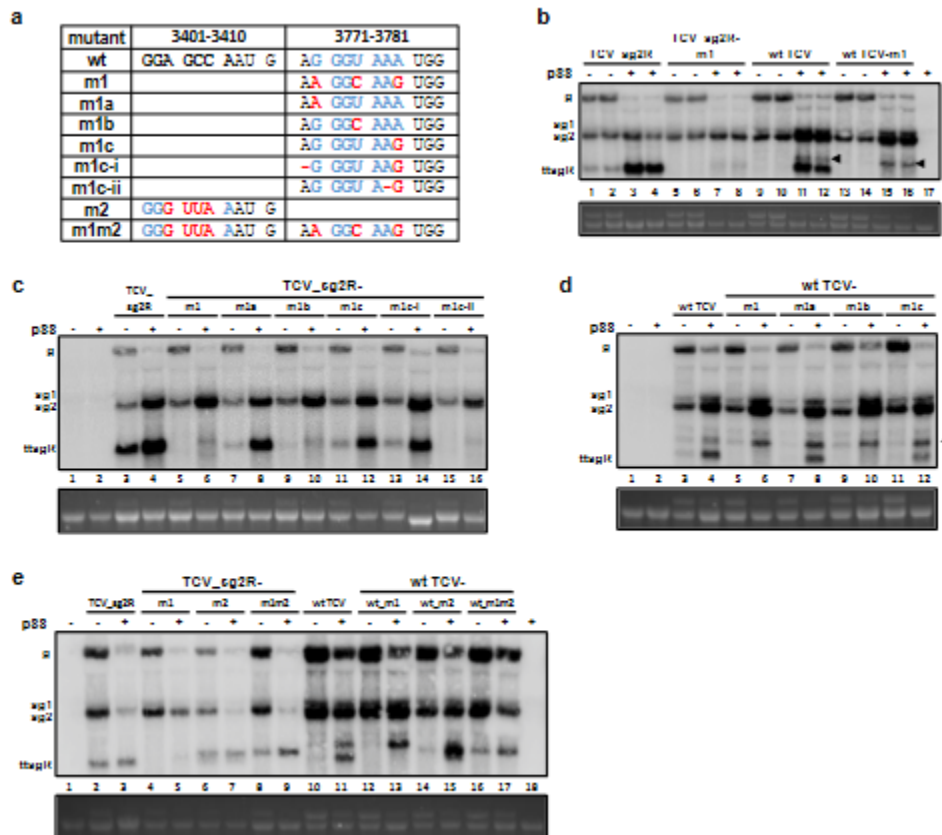


**Figure 7**. The $G_3(A/U)_4$ motif is necessary and sufficient to initiate the synthesis of ttsgR and similar TCV lncRNAs. **a**. wt TCV sequences at positions of nt #3,401-3410, and nt #3,771-3,781, and sequences of various mutants that altered one or both of these regions. The sequences

are separated by spaces into codons in CP ORF. The wildtype GGGUAAA motif responsible for

ttsgR initiation is highlighted blue. All mutated nt are highlighted in red. **b, c, d,** and **e**. Northern

blots showing the impact of various mutations on the levels of ttsgR and related lncRNAs

initiated from replicons. **b**. TCV_sg2R-m1 and wt TCV-m1, both containing the same three

single-nt mutations within the GGGUAAA motif, suffered a near complete loss of ttsgR, with or

without p88. The two arrowheads denote an additional lncRNA present in wt TCV lanes, which

could have initiated from a $G_3(A/U)_4$ motif within the TCV CP region that was by mCherry (nt

#3,290-3,296, GGGUUUA) in TCV_sg2R. **c**. The impact of several single nt mutations (m1a,

m1b, m1c, m1c-i, m1c-ii) on the level of ttsgR in the TCV_sg2R background. **d**. The impact of

m1a, m1b, and m1c mutations on the level of ttsgR in the wt TCV background. Note the

arrowhead denoting the position of the additional lncRNA unique to wt TCV. **e**. Creation of a

new $G_3(A/U)_4$ motif (m2, GGGUUAA at nt #3,401-3,407) was sufficient to program the

production of a new lncRNA, in both TCV_sg2R and wt TCV backgrounds.


We also verified the identity of the new lncRNA by analyzing the progeny sequences

obtained from *N. benthamiana* cells infected with the m1m2 mutant (in TCV_sg2R background).

As shown in S3 Fig, more than half of the clones (14 out of 23) contained lncRNA sequences

that initiated at the new $G_3(A/U)_4$ motif, whereas none of them initiated from the original site

where the $G_3(A/U)_4$ motif was abolished. Interestingly, we again recovered sequences that

initiated at nt #3,802 (7 out of 23), verifying the frequent usage of this GGUAAU site as an

initiation point for lncRNA. Similarly interesting were 2 sequences that initiated from nt #3,740,

though in the resulting lncRNA the GUGAAAA motif was changed into GGGAAAA. Together

these results demonstrated that a $G_3(A/U)_4$ motif was sufficient to launch the synthesis of a new

lncRNA (or sgRNA) from an internal position of TCV gRNA, whose accumulation were bolstered by p88 overexpression.

**The m1 mutant of wt TCV retains high infectivity under ordinary conditions**. Finally we wondered whether production of ttsgR was critical to efficient systemic infection of TCV in host plants. To this end, we used the m1 mutant of wt TCV to infect *N. benthamiana* plants, and closely monitored the infected plants for six weeks. Compared to plants infected the wt TCV control, these plants did not show any significant differences in the timing and severity of symptoms. The titers of viral RNA as determined for multiple plants, using Northern blotting, were very similar as well. Furthermore, the m1 mutant virus remained stable for at least four sequential passages in *N. benthamiana*, and continued to cause symptoms indistinguishable from wt TCV. We also carried out competition experiments by infecting *N. benthamiana* plants with a mixture of m1 and wt TCV. Both variants were recovered at similar frequencies in the systemically infected leaves of six mix-infected plants (Table 1). Indeed among a total of 26 clones sequenced, 17 were m1, 9 were wt TCV, suggesting that the m1 mutant was not competitively disadvantaged. Therefore, diminished ttsgR production did not appear to have a measurable impact on TCV infectivity in this particular host, under the growth condition we used. However, it is possible that ttsgR might play a role in TCV infectivity in a different host plant, or when the plants are under various stresses. On the other hand, other lncRNAs initiating from alternative $G_3(A/U)_4$ motifs (e.g. nt #3,740, 3,802, or 3290) might compensate for the loss of ttsgR.

**Table 1.** Persistence of the m1 mutant of TCV in plants mix-infected with wt TCV

| Plant # | # of clones with AGGCAAG (m1) | # of clones with GGGUAAA (wt) |
|---------|-------------------------------|-------------------------------|
| 1 | 3 | 0 |
| 2 | 4 | 1 |

| 3 | 1 | 3 |
|---|---|---|
| 4 | 2 | 3 |
| 5 | 3 | 2 |
| 6 | 4 | 0 |
| **Total** | **17** | **9** |

**Discussion**

**Replication-based production of ttsgR.** Viral lncRNAs have been reported for numerous (+) RNA viruses. They have been well investigated for a number of human-infecting flaviviruses, including DENV, WNV, and Zika virus, which are collectively known as long noncoding subgenomic flavivirus RNAs (sfRNAs) [16,28,50]. Viral lncRNAs have also been widely observed in (+) RNA plant virus infections, including BNYVV, RCNMV, and TNV-D [17,19,22]. Most of these viral lncRNAs share the following characteristics: (i) they are typically less than 500 nt in length, and often heterogenous at the 5' end; (ii) they are derived from the 3' termini of parental viruses; (iii) they are produced by exonuclease degradation of longer viral RNAs that stalls at various stable secondary structures; and (iv) they mostly function to benefit the parental viruses.

ttsgR shares most of these characteristics but is notably different in one key aspect: unlike the previously examined viral lncRNAs, ttsgR is produced by the viral replication process. This conclusion is well supported by several lines of evidence. First, ttsgR was readily produced from a template that was the same size as the final product, in the absence of any other longer viral RNA, but in the presence of viral replication proteins p28 and p88. Second, p28 and/or p88-mediated template stabilization was unlikely the cause for ttsgR accumulation because replacing p88 with the RdRp-null p88mGDD completely abolished it. Third, both (+) and (-) strands of ttsgR were detectable, suggesting that the (+) template was copied into (-) intermediates through

replication. Fourth, dimeric forms of ttsgR were readily detectable, which could only have been produced through inaccurate copying, possibly on non-covalently circularized templates [51,52]. Finally, production of ttsgR depended on the presence of the CCC motif at the 3' end, and a $G_3(A/U)_4$ motif at the 5' end. The CCC motif was previously established by others as essential for the replication of TCV gRNA and a satellite RNA associated with TCV (satC) [47,53]. Similarly, the $G_3(A/U)_4$ motif was also found to be required for the (-)-to-(+) strand replication of satC, and widely conserved among carmovirus genomic and subgenomic RNAs [48,49]. Together these findings implicate the TCV replication machinery in the biogenesis of ttsgR, and unveil the first replicating viral lncRNA.

While replication-based production has not been reported for other viral lncRNAs, this production mode should not be surprising given the functional conservation of viral lncRNAs. Namely, if these lncRNAs play important pro-viral roles during infections, the mechanisms of their biogenesis can be expected to evolve independently, converging on similar end products. Another important point we wish to make is that none of the previously reported viral lncRNAs have been examined using an *in vivo* replication system that is independent of the replicating parental virus. It may not be unthinkable that some of those earlier lncRNAs could be produced/stabilized by a combination of replication and stalled degradation. Indeed, a recent report found that the (-) strand form of RCNMV lncRNA was detectable in infected tissues, suggesting that this lncRNA may have resulted from such a combination [54]. Separately, one of the lncRNA species derived from Japanese encephalitis virus (JEV) was found to act as a potent template for JEV RdRp *in vitro* [55].

**Role of p88 overexpression in ttsgR.** It is worth noting that ttsgR accumulation was rather low in infections with wt TCV or TCV replicons that encode their own p28 and p88. However,

its levels were dramatically elevated by p88 overexpression. We reported earlier that moderate p88 overexpression depressed the accumulation of TCV gRNA, but elevated the accumulation of sgRNAs, especially sgRNA2 [38]. The fact that both sgRNA2 and ttsgR were preferentially amplified by p88 overexpression hints at a regulative valve controlled by p88 protein levels. It is possible that higher p88 levels, which could occur at a late stage of TCV cellular infections, serves as a switch that down-regulates gRNA replication in favor of sgRNA and ttsgR replication (transcription). How can this be achieved? We speculate that higher p88 levels may lower the p28/p88 ratio, favoring the formation of smaller membrane-encircled spherules in which RNA replication occur [56]. It is further possible that such smaller spherules act as constraints to limit the size of the RNA templates they enclose, leading to preferential amplification of smaller RNA species.

**Conservation and repair of ttsgR termini.** Since ttsgR is produced through replication, it is no surprise that its termini share sequence/structure characteristics with TCV gRNA and sgRNAs. Accordingly, the CCC motif at the 3' terminus was found to be essential for ttsgR accumulation, as altering this motif by as few as one nt led to dramatic diminution, or complete loss of ttsgR accumulation. Furthermore, the mutations at the middle C invariably reverted to the original identity in ttsgR variants recovered from the poorly accumulating pools (Zhang and Qu, unpublished). Therefore, the repair (reversion) of the 3' terminus, possibly occurring during (-) strand synthesis, was very inefficient, conceivably occurring in only a fraction of cells. Successful repair of deleted 3' ends have been reported for TCV-associated satellite RNAs[47,53]. However, those repairs occurred in the presence of a helper TCV RNA that donated the repair templates.

By contrast, mutations at the 5' end of ttsgR, within the $G_3(A/U)_4$ motif, were very efficiently repaired, suggesting that altering the 5' end had a relatively minor impact on the synthesis of (-) strands, and were instead repaired during the process of synthesizing (+) strands from (-) strand intermediates. This conclusion is consistent with previous studies showing that a complementary motif at the 3' end of the satC (-) strand, $CC_{1-2}(A/U)(A/U)(A/U)$, was important for the synthesis of (+) strand satC [49]. Interestingly, the 5' end repair was very flexible, accomplished through either the use of alternative starting sites, or *de novo* addition of extra nt. Nevertheless, most successful variants always restored characteristic $G_3(A/U)_4$ motif. As previously noted, this 5' end motif is widely conserved among genomic and subgenomic RNAs of most carmoviruses [48,49].

Importantly, $G_3(A/U)_4$ motifs present in the TCV gRNA internally, especially those within its 3' half, drove potent initiation of viral lncRNAs that over-accumulated in the presence of excess RdRp. In addition to the GGGUAAA motif at nt #3,772 responsible for ttsgR accumulation, similar motifs at nt #3,740 (GUGAAAA) and #3,803 (GGUAAU) likewise programed the synthesis of lncRNAs, albeit at lower abundance. A fourth motif at nt #3,290 (GGGUUUA) likely caused the initiation of another yet-to-be characterized lncRNA seen in multiple Northern blots. Finally, creating a GGGUUAA motif at a new location (nt #3,401) in TCV gRNA was enough to program the production of a novel lncRNA. Together these observations reinforce the critical role of the $G_3(A/U)_4$ motif in the replication of diverse TCV RNAs, including gRNA and both sgRNAs.

Unlike sgRNAs, replication of ttsgR and similar lncRNAs did not require strong secondary structures prior to their initiation sites [42,43]. These strong RNA secondary structures, known as attenuation structures, cause premature termination of (-) strand synthesis at uniform locations,

thus ensuring precise 3' termini for (-) strand sgRNAs, which coincide with the $G_3(A/U)_4$ motif

at the 5' end of (+) strands. However, spontaneous premature termination of (-) strand synthesis

could occur at random positions with low frequencies. With the presence of overexpressed p88

and the smaller replication-enabling spherules (discussed above), some of the spontaneously

terminated (-) strands could become protected from degradation, and serve as templates for

lncRNA synthesis. Nevertheless, the presence of the $G_3(A/U)_4$ motif at the 5' termini of all

successful lncRNAs suggests that this motif is a key recognition signal of p88 RdRp. Indeed, it

appears that p88 might use the same CCC motif to template the synthesis of both (-) and (+)

strands of TCV RNAs.

   **Potential interactions between ttsgR (and other TCV lncRNAs) and TCV-associated**

**satellite RNAs.** The small size of ttsgR (283 nt), along with its characteristic 5' end (at nt #3,772

of TCV gRNA), prompts the speculation that ttsgR or other similar lncRNAs might have played

a role in the birth of satC and its variants. satC is an intriguing TCV-parasitic satellite RNA that

contains both TCV-derived sequences and that of satD, another TCV-parasitic satellite RNA.

Specifically, the 356-nt satC consists of a 189-nt segment derived from satD at its 5' end, a 16-nt

(#3,764-3,779) TCV fragment in the middle, and the last 151-nt (#3,900-4,054) of TCV gRNA at

the 3' end [57]. Interestingly, the first 8 nt of ttsgR corresponds to the second half of the 16-nt

middle section of satC (#3772-3779), suggesting that satC might have derived from

recombination events between satD and ttsgR-like lncRNAs, followed by deletion of nt #3,780-

3899 of TCV gRNA.

   Given this possibility, we were surprised to find that ttsgR does not qualify as a defective

interfering (DI) RNA, because a full-length TCV replicon (e.g. TCV_sg2R or wt TCV) failed to

replicate ttsgR from a separate template (as opposed to from the replicons themselves) (Sun and

Qu, unpublished). This is despite the fact that p28 and p88 provided from non-viral sources

readily replicated ttsgR from such a template. These contrasting observations await further

examinations to identify additional elements that convert ttsgR into a DI. In summary, the 283 nt

ttsgR offers a dramatically simplified substrate on which many aspects of the TCV replication

process can be examined.

In contrast to previously reported viral lncRNAs that were found to have diverse pro-virus

and anti-defense roles, we have yet to delineate a definitive role for ttsgR. However, it is possible

that ttsgR confers robustness to TCV under certain yet-to-be tested conditions that threat viral

persistence. Alternatively, TCV lncRNAs initiating at other positions (e.g. nt #3,802, #3,740, and

possibly #3,290) could function to compensate the loss of ttsgR. In summary, we have

discovered and characterized a novel viral lncRNA that is produced by viral replication. Further

dissection of this and other TCV lncRNAs is expected to clarify their potential functions and

offer new targets for virus control and management.


**Material and Methods**

**Constructs**. Binary constructs that express TCV p28 and TBSV p19 from the strong

2X35S promoter were generated earlier [35,39]. Those harboring TCV replicons, including

TCV_sg2R, Δp88_sg2R, [p28stop]_sg2R, as well as wt TCV, were also described in earlier

studies [35,38,39]. Constructs designed to express various forms of p88, all of them harboring a

C-terminal duplicated HA tag, including 2X35S::p88, Core::p88, Core::p88ΔN36,

Core::p88ΔN127, were generated in a recent study [38]. A new p88-expressing construct,

Micro::p88, were produced by replacing the Core 35S promoter of the Core::p88 construct with a

microspore-specific promoter derived from *Nicotiana tabacum* [41]. The p88mGDD construct

was produced by replacing the HpaI-SalI fragment in the p88 coding region of the Core::p88

construct, with a synthesized DNA fragment (gBlock) in which the sequence encoding the GDD

motif (GGAGACGAT) was altered to encode VAA (G<u>TT</u>GC<u>A</u>GC<u>T</u>) (S1 Table. Integrated DNA

Technology, Coralville, IA). The replacement was accomplished using Gibson Assembly

ligation (New England Biolabs, Ipswich, MA).

Mutant replicon constructs Δsg1, Δsg2, and Δsg1&2, all of them derivatives of TCV_sg2R,

were similarly created with synthesized DNA fragments (gBlocks) and Gibson Assembly

cloning. Specifically, the gBlock for Δsg1 contained 5 single nt substitutions known to disrupt

the stem-loop structure of the sgRNA1 promoter [42](Fig. 2A. S1 Table). It was used to replace

the SalI-BamHI fragment in TCV_sg2R. The gBlock for Δsg2 contained a 12 nt deletion that

disrupts the stem-loop structure of the sgRNA2 promoter [42,43] (Fig. 2A. S1 Table). It was

used to replace the BamHI-AatII fragment in TCV_sg2R. The Δsg1 and Δsg2 gBlocks were

combined in a three-fragment Gibson Assembly to replace the SalI-AatII fragment of

TCV_sg2R, creating the Δsg1&2 mutant.

The m1, m1a, m1b, m1c, m2, and m1m2 mutants, in both TCV_sg2R and wt TCV

backbones, were created by using PCR-generated overlapping DNA fragments that were

subsequently incorporated into TCV_sg2R or wt TCV digested with AatII (located at nt #2,600

of TCV cDNA) plus SpeI (nt #3,950 of TCV cDNA) using multi-fragment Gibson Assembly.

The oligo-deoxyribonucleotide primers used in these PCR reactions are provided in S2 Table.

The construct serving as the template for sgRNA2, sg2R_Temp, was made by removing the

first 2,600 nt of the TCV_sg2R insert with XhoI (located at the extreme 5' end of TCV cDNA)

plus AatII digestions, followed by refilling the gap with a gBlock fragment spanning nt #2468-

2600 (S2 Table), thus restoring the stem-loop structure of the sgRNA2 promoter. The template

constructs ttsgR_Temp and ttsgR_S were produced by replacing the XhoI-SpeI fragment of TCV_sg2R with gBlock fragments encompassing nt #3386-3950, and 3744-3950 of TCV gRNA (S1 Table), respectively. Constructs ttsgR_mL, ttsgR_mR, and ttsgR_mLR are variants of ttsgR_Temp, each made with gBlock fragments that contained mutations shown in Fig. 2D. Finally, the ttsgR_S2 and ttsgR_S2a templates, as well as mutants thereof, were generated using PCR, with primers listed in S2 Table. The identities of all new constructs were verified with Sanger sequencing.

**Additional Materials and Methods** are provided in the Supplementary Texts and Tables.

### References

1.  Flint SJ, Racaniello VR, Rall GF, Skalka AM, Enquist LW. Principles of virology. 2015.

2.  Razum O, Sridhar D, Jahn A, Zaidi S, Ooms G, Müller O. Polio: from eradication to systematic, sustained control. BMJ Global Health. 2019;4: e001633. doi:10.1136/bmjgh-2019-001633

3.  Martina BEE, Koraka P, Osterhaus ADME. Dengue Virus Pathogenesis: an Integrated View. Clin Microbiol Rev. 2009;22: 564. doi:10.1128/CMR.00035-09

4.  Petersen LR, Jamieson DJ, Powers AM, Honein MA. Zika Virus. N Engl J Med. 2016;374: 1552–1563. doi:10.1056/NEJMra1602113

5.    Cevik M, Kuppalli K, Kindrachuk J, Peiris M. Virology, transmission, and pathogenesis of SARS-CoV-2. BMJ. 2020;371: m3862. doi:10.1136/bmj.m3862

6.    Palazzo AF, Koonin EV. Functional Long Non-coding RNAs Evolve from Junk Transcripts. Cell. 2020;183: 1151–1161. doi:10.1016/j.cell.2020.09.047

7.    Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nature Reviews Genetics. 2016;17: 47–62. doi:10.1038/nrg.2015.10

8.    Scholthof HB, Jackson AO. The Enigma of pX: A Host-Dependentcis-Acting Element with Variable Effects on Tombusvirus RNA Accumulation. Virology. 1997;237: 56–65. doi:10.1006/viro.1997.8754

9.    Kelly L, Gerlach WL, Waterhouse PM. Characterisation of the Subgenomic RNAs of an Australian Isolate of Barley Yellow Dwarf Luteovirus. Virology. 1994;202: 565–573. doi:10.1006/viro.1994.1378

10.   Koev G, Miller WA. A Positive-Strand RNA Virus with Three Very Different Subgenomic RNA Promoters. J Virol. 2000;74: 5988. doi:10.1128/JVI.74.13.5988-5996.2000

11.   Balmori E, Gilmer D, Richards K, Guilley H, Jonard G. Mapping the promoter for subgenomic RNA synthesis on beet necrotic yellow vein virus RNA 3. Biochimie. 1993;75: 517–521. doi:10.1016/0300-9084(93)90056-X

12.   Scheets K. Maize Chlorotic Mottle Machlomovirus Expresses Its Coat Protein from a 1.47-kb Subgenomic RNA and Makes a 0.34-kb Subgenomic RNA. Virology. 2000;267: 90–101. doi:10.1006/viro.1999.0107

13.   Urosevic N, van Maanen M, Mansfield JP, Mackenzie JS, Shellam GR. Molecular characterization of virus-specific RNA produced in the brains of flavivirus-susceptible and -resistant mice after challenge with Murray Valley encephalitis virus. Journal of General Virology. Microbiology Society; 1997. pp. 23–29. doi:https://doi.org/10.1099/0022-1317-78-1-23

14.   Lin K-C, Chang H-L, Chang R-Y. Accumulation of a 3'-terminal genome fragment in Japanese encephalitis virus-infected mammalian and mosquito cells. J Virol. 2004;78: 5133–5138. doi:10.1128/jvi.78.10.5133-5138.2004

15.   Scherbik SV, Paranjape JM, Stockman BM, Silverman RH, Brinton MA. RNase L Plays a Role in the Antiviral Response to West Nile Virus. J Virol. 2006;80: 2987. doi:10.1128/JVI.80.6.2987-2999.2006

16.   Pijlman GP, Funk A, Kondratieva N, Leung J, Torres S, van der Aa L, et al. A Highly Structured, Nuclease-Resistant, Noncoding RNA Produced by Flaviviruses Is Required for Pathogenicity. Cell Host & Microbe. 2008;4: 579–591. doi:10.1016/j.chom.2008.10.007

17.   Iwakawa H, Mizumoto H, Nagano H, Imoto Y, Takigawa K, Sarawaneeyaruk S, et al. A Viral Noncoding RNA Generated by *cis*-Element-Mediated Protection against 5'→3' RNA Decay Represses both Cap-Independent and Cap-Dependent Translation. J Virol. 2008;82: 10162. doi:10.1128/JVI.01027-08

18. Pallarés HM, Costa Navarro GS, Villordo SM, Merwaiss F, de Borba L, Gonzalez Lopez Ledesma MM, et al. Zika Virus Subgenomic Flavivirus RNA Generation Requires Cooperativity between Duplicated RNA Structures That Are Essential for Productive Infection in Human Cells. López S, editor. J Virol. 2020;94: e00343-20. doi:10.1128/JVI.00343-20

19. Gunawardene CD, Newburn LR, White KA. A 212-nt long RNA structure in the Tobacco necrosis virus-D RNA genome is resistant to Xrn degradation. Nucleic Acids Research. 2019;47: 9329–9342. doi:10.1093/nar/gkz668

20. Dilweg IW, Gultyaev AP, Olsthoorn RC. Structural features of an Xrn1-resistant plant virus RNA. RNA Biol. 2019/04/05 ed. 2019;16: 838–845. doi:10.1080/15476286.2019.1592070

21. Flobinus A, Chevigny N, Charley P, Seissler T, Klein E, Bleykasten-Grosshans C, et al. Beet Necrotic Yellow Vein Virus Noncoding RNA Production Depends on a 5'→3' Xrn Exoribonuclease Activity. Viruses. 2018;10: 137. doi:10.3390/v10030137

22. Peltier C, Klein E, Hleibieh K, D'Alonzo M, Hammann P, Bouzoubaa S, et al. Beet necrotic yellow vein virus subgenomic RNA3 is a cleavage product leading to stable non-coding RNA required for long-distance movement. Journal of General Virology. : 10.

23. Miller WA, Shen R, Staplin W, Kanodia P. Noncoding RNAs of Plant Viruses and Viroids: Sponges of Host Translation and RNA Interference Machinery. : 9.

24. MacFadden A, O'Donoghue Z, Silva PAGC, Chapman EG, Olsthoorn RC, Sterken MG, et al. Mechanism and structural diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs. Nat Commun. 2018;9: 119. doi:10.1038/s41467-017-02604-y

25. Funk A, Truong K, Nagasaki T, Torres S, Floden N, Balmori Melian E, et al. RNA Structures Required for Production of Subgenomic Flavivirus RNA. JVI. 2010;84: 11407–11417. doi:10.1128/JVI.01159-10

26. Silva PAGC, Pereira CF, Dalebout TJ, Spaan WJM, Bredenbeek PJ. An RNA Pseudoknot Is Required for Production of Yellow Fever Virus Subgenomic RNA by the Host Nuclease XRN1. JVI. 2010;84: 11395–11406. doi:10.1128/JVI.01047-10

27. Steckelberg A-L, Vicens Q, Kieft JS. Exoribonuclease-Resistant RNAs Exist within both Coding and Noncoding Subgenomic RNAs. Idnurm A, editor. mBio. 2018;9: e02461-18, /mbio/9/6/mBio.02461-18.atom. doi:10.1128/mBio.02461-18

28. Chapman EG, Costantino DA, Rabe JL, Moon SL, Wilusz J, Nix JC, et al. The Structural Basis of Pathogenic Subgenomic Flavivirus RNA (sfRNA) Production. Science. 2014;344: 307–310. doi:10.1126/science.1250897

29. Chapman EG, Moon SL, Wilusz J, Kieft JS. RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. eLife. 2014;3: e01892. doi:10.7554/eLife.01892

30. Akiyama BM, Laurence HM, Massey AR, Costantino DA, Xie X, Yang Y, et al. Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. : 6.

31.  Steckelberg A-L, Akiyama BM, Costantino DA, Sit TL, Nix JC, Kieft JS. A folded viral noncoding RNA blocks host cell exoribonucleases through a conformationally dynamic RNA structure. Proc Natl Acad Sci USA. 2018;115: 6404–6409. doi:10.1073/pnas.1802429115

32.  Steckelberg A-L, Vicens Q, Costantino DA, Nix JC, Kieft JS. The crystal structure of a Polerovirus exoribonuclease-resistant RNA shows how diverse sequences are integrated into a conserved fold. : 11.

33.  Slonchak A. Subgenomic flaviviral RNAs_ What do we know after the first decade of research. Antiviral Research. 2018; 13.

34.  Cao M, Ye X, Willie K, Lin J, Zhang X, Redinbaugh MG, et al. The Capsid Protein of Turnip Crinkle Virus Overcomes Two Separate Defense Barriers To Facilitate Systemic Movement of the Virus in Arabidopsis. Journal of Virology. 2010;84: 7793–7802. doi:10.1128/JVI.02643-09

35.  Qu F, Ren T, Morris TJ. The Coat Protein of Turnip Crinkle Virus Suppresses Posttranscriptional Gene Silencing at an Early Initiation Step. Journal of Virology. 2003;77: 511–522. doi:10.1128/JVI.77.1.511-522.2003

36.  Qu F, Ye X, Hou G, Sato S, Clemente TE, Morris TJ. RDR6 Has a Broad-Spectrum but Temperature-Dependent Antiviral Defense Role in Nicotiana benthamiana. Journal of Virology. 2005;79: 15209–15217. doi:10.1128/JVI.79.24.15209-15217.2005

37.  Qu F, Ye X, Morris TJ. Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. Proceedings of the National Academy of Sciences. 2008;105: 14732–14737.

38.  Zhang S, Sun R, Guo Q, Zhang X-F, Qu F. Repression of turnip crinkle virus replication by its replication protein p88. Virology. 2019;526: 165–172. doi:10.1016/j.virol.2018.10.024

39.  Zhang X-F, Sun R, Guo Q, Zhang S, Meulia T, Halfmann R, et al. A self-perpetuating repressive state of a viral replication protein blocks superinfection by the same virus. PLOS Pathogens. 2017;13: e1006253. doi:10.1371/journal.ppat.1006253

40.  Guo Q, Zhang S, Sun R, Yao X, Zhang X-F, Tatineni S, et al. Superinfection Exclusion by p28 of Turnip Crinkle Virus Is Separable from Its Replication Function. MPMI. 2020;33: 364–375. doi:10.1094/MPMI-09-19-0258-R

41.  Oldenhof MT, de Groot PFM, Visser JH, Schrauwen JAM, Wullems GJ. Isolation and characterization of a microspore-specific gene from tobacco. Plant Molecular Biology. 1996;31: 213–225. doi:10.1007/BF00021785

42.  Wang J, Simon AE. Analysis of the Two Subgenomic RNA Promoters for Turnip Crinkle Virusin Vivoandin Vitro. Virology. 1997;232: 174–186. doi:10.1006/viro.1997.8550

43.  Wu B, Oliveri S, Mandic J, White KA. Evidence for a Premature Termination Mechanism of Subgenomic mRNA Transcription in a Carmovirus. J Virol. 2010;84: 7904. doi:10.1128/JVI.00742-10

44.   Shivaprasad PV, Akbergenov R, Trinks D, Rajeswaran R, Veluthambi K, Hohn T, et al. Promoters, Transcripts, and Regulatory Proteins of Mungbean Yellow Mosaic Geminivirus. J Virol. 2005;79: 8149–8163. doi:10.1128/JVI.79.13.8149-8163.2005

45.   Plaskon NE, Adelman ZN, Myles KM. Accurate Strand-Specific Quantification of Viral RNA. PLOS ONE. 2009;4: e7468. doi:10.1371/journal.pone.0007468

46.   Qu F, Morris TJ. Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. J Virol. 1997;71: 1428.

47.   Nagy PD, Carpenter CD, Simon AE. A novel 3'-end repair mechanism in an RNA virus. Proc Natl Acad Sci USA. 1997;94: 1113. doi:10.1073/pnas.94.4.1113

48.   Guan H, Song C, Simon AE. RNA promoters located on (-)-strands of a subviral RNA associated with turnip crinkle virus. RNA. 1997;3: 1401–1412.

49.   Guan H, Carpenter CD, Simon AE. Analysis of cis-Acting Sequences Involved in Plus-Strand Synthesis of a Turnip Crinkle Virus-Associated Satellite RNA Identifies a New Carmovirus Replication Element. Virology. 2000;268: 345–354. doi:10.1006/viro.1999.0153

50.   Zhang Y, Zhang Y, Liu Z-Y, Cheng M-L, Ma J, Wang Y, et al. Long non-coding subgenomic flavivirus RNAs have extended 3D structures and are flexible in solution. EMBO reports. 2019;20: e47016. doi:10.15252/embr.201847016

51.   Herold J, Andino R. Poliovirus RNA Replication Requires Genome Circularization through a Protein–Protein Bridge. Molecular Cell. 2001;7: 581–591. doi:10.1016/S1097-2765(01)00205-2

52.   Li Z, Nagy PD. Diverse roles of host RNA binding proteins in RNA virus replication. RNA Biology. 2011;8: 305–315. doi:10.4161/rna.8.2.15391

53.   Carpenter CD, Simon AE. In vivo restoration of biologically active 3' ends of virus-associated RNAs by nonhomologous RNA recombination and replacement of a terminal motif. J Virol. 1996;70: 478.

54.   Newburn LR, White KA. A trans-activator-like structure in RCNMV RNA1 evokes the origin of the trans-activator in RNA2. Wang A, editor. PLoS Pathog. 2020;16: e1008271. doi:10.1371/journal.ppat.1008271

55.   Chen Y-S, Fan Y-H, Tien C-F, Yueh A, Chang R-Y. The conserved stem-loop II structure at the 3' untranslated region of Japanese encephalitis virus genome is required for the formation of subgenomic flaviviral RNA. PLOS ONE. 2018;13: e0201250. doi:10.1371/journal.pone.0201250

56.   Ertel KJ, Benefield D, Castaño-Diez D, Pennington JG, Horswill M, den Boon JA, et al. Cryo-electron tomography reveals novel features of a viral RNA replication compartment. Elife. 2017;6. doi: 10.7554/eLife.25940

57.   Simon AE, Howell SH. The virulent satellite RNA of turnip crinkle virus has a major domain homologous to the 3' end of the helper virus genome. The EMBO Journal. 1986;5: 3423–3428. doi:10.1002/j.1460-2075.1986.tb04664.x