Unusual characteristics of dicistrovirus-derived small RNAs in the small brown planthopper, *Laodelphax striatellus*

Junmin Li,†,1 Ida Bagus Andika,† Yanru Zhou,2 Jiangfeng Shen,2 Zongtao Sun,† Xu Wang,† Liying Sun† and Jianping Chen†

1State Key Laboratory Breeding Base for Zhejiang Sustainable Pest and Disease Control, Ministry of Agriculture Key Laboratory of Biotechnology in Plant Protection, Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, PR China
2College of Chemistry and Life Science, Zhejiang Normal University, Jinhua 321004, PR China

In this study, sequences of small RNA (sRNA) libraries derived from the insect vector *Laodelphax striatellus* were assembled into contigs and used as queries for database searches. A large number of contigs were highly homologous to the genome sequence of an insect dicistrovirus, himetobi P virus (HiPV). Interestingly, HiPV-derived sRNAs had a wide size distribution, and were relatively abundant throughout the 18–30 nt size range with only a slight peak at 22 nt. HiPV sRNAs had a strong bias towards the sense strand, whilst the antisense sRNAs were predominantly 21 and 22 nt. HiPV sRNAs do not have the typical features of PIWI-interacting RNAs, but their 3’ ends were preferentially cleaved at UA-rich sequences. Our data suggest that HiPV sRNAs may be derived both from activities of the RNA interference pathway and from cleavage of the viral genome by other host RNases.

The antiviral RNA interference (RNAi) pathway limits virus accumulation in cells through the generation of virus-derived small interfering RNAs (siRNAs) that guide the sequence-specific degradation of the viral RNA target (Ding & Voinnet, 2007). In recent years, deep sequencing of small RNAs (sRNAs) has been used to comprehensively characterize virus-derived siRNAs in various hosts (Aregger et al., 2012; Ding & Lu, 2011; Li et al., 2013; Siu et al., 2011; Xu et al., 2012). This method is able to identify large numbers of viral siRNAs that overlap in sequence with each other along the viral genome and make it possible to assemble them into longer contiguous sequences (contigs). Several studies have demonstrated that when contigs derived from the assembly of sRNA sequences are used as queries for database searches, many contigs are homologous to the genome sequences of known viruses, including some not known to infect those hosts (Bi et al., 2012; Kreuze et al., 2009; Ma et al., 2011; Sela et al., 2012; Wu et al., 2010). Hence, sRNA deep sequencing is a powerful and cost-effective approach for providing a global view of virus populations within a host and also for the discovery of new viruses (Liu et al., 2011; Singh et al., 2012).

Previously, we characterized siRNAs derived from rice black-streaked dwarf virus (RBSDV; genus *Fijivirus*, family *Reoviridae*) and rice stripe virus (RSV; genus *Tenuivirus*, family unassigned) in the insect vector *Laodelphax striatellus* (small brown planthopper) by deep sequencing (Li et al., 2013). RBSDV and RSV are rice-infecting viruses that are also able to replicate in *L. striatellus* (Shinkai, 1962; Toriyama, 1986). RBSDV and RSV siRNAs were predominantly 21 and 22 nt, and there were almost equal proportions of positive- and negative-sense molecules (Li et al., 2013). In that study, we also established co-infection of RBSDV and RSV in *L. striatellus* under laboratory conditions, providing the first report demonstrating simultaneous infection by different viruses in this insect. It is known that planthoppers can transmit several economically important rice viruses that can all replicate in insect cells (Hogenhout et al., 2008), and it is therefore possible that *L. striatellus* can be simultaneously infected by multiple viruses under natural conditions.

In our previous study, we obtained four sRNA libraries derived from *L. striatellus* that were infected with RBSDV (SI_RB library) or RSV (SI_RSV library), doubly infected with RBSDV and RSV (DI library), or not infected with either of these viruses (VF library). Taking advantage of the availability of these libraries, we wanted to investigate...
whether our laboratory *L. striatellus* cultures were also infected with other known or unknown viruses. After the removal of RBSDV and RSV siRNAs, the unique sRNA sequences of each library were assembled *de novo* using the program Velvet, version 1.1 (Zerbino & Birney, 2008), with the parameters; k-mer 18 for velveth program, cov_cut-off auto, min_contig_lgth 50 for velvetg program and default for other parameters. The assembly yielded approximately 2500 contigs (>50 nt) with various lengths from each library. When the assembled contigs were used as queries for BLASTN searches against the viral genomes database available in the NCBI database (NCBI RefSeq, ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral/), a large number of contigs derived from each of the four libraries showed high identities (e-value <10^-10) to himetobi P virus (HiPV) (Table S1). HiPV belongs to the genus Cripavirus in the family Dicistroviridae and was previously reported to be isolated from *L. striatellus* and the brown planthopper, *Nilaparvata lugens* (Guy et al., 1992; Toriyama et al., 1992). The HiPV genome is a single positive-stranded RNA (~9.2 kb), and contains two large ORFs encoding a replication protein and a capsid protein precursor (Nakashima et al., 1999) (see also Fig. 1a).

To confirm the infection of HiPV in our laboratory *L. striatellus* cultures, strand-specific reverse transcription-PCR (RT-PCR) was carried out using reverse primer HiPV-R (5’-ACCAAAGAGTGCCACAGGAG-3’) and forward primer HiPV-F (5’-GTTGTTAACCGCAGCACGTA-3’) specific for the 5’-terminal untranslated region (UTR) of the virus genome. For detection of positive- and negative-strand genomic RNA, the reverse and forward primers were used, respectively, in the reverse transcription step. Total RNA was extracted from individual insects using a method described previously (Li et al., 2013). Both positive and negative strands of the HiPV genome were detected by RT-PCR in all insects examined (*n*=100) (Fig. 1b and data not shown). Amplification of HiPV sequence was verified by sequencing of the RT-PCR products. This result indicated a high infection rate of HiPV in our insect cultures. Furthermore, detection of the negative-strand RNA genome by RT-PCR confirmed the replication of HiPV in our *L. striatellus* cultures. Quantitative RT-PCR analysis of HiPV RNA in the four different samples showed that the accumulation level was much higher in SI_RSV than DI or the two other samples (SI_RB and VF) (Fig. 1c), suggesting that co-infection with RSV but not with RBSDV enhanced HiPV accumulation.

HiPV homologous contigs assembled from the sRNA libraries covered the entire genome of previously reported HiPV (Guy et al., 1992) (GenBank accession no. NC_003782.1) and could therefore be assembled into a full genome using DNAMAN 6 (Fig. 1a). The sequences of the RT-PCR products amplified using primers specific for the 5’ UTR, replicase- and capsid-coding regions were identical to the sequences assembled from the viral sRNA contigs (data not shown), indicating the reliability of the assembly. As the *L. striatellus* cultures used in this study originated from Jiangsu Province, China, the newly

---

**Fig. 1.** Identification of HiPV infection in *L. striatellus* by assembly of small RNAs. (a) Mapping of contigs (represented by black lines) derived from the assembly of *L. striatellus* sRNA sequences (VF library) to the genome of HiPV. Previously, the sequence homology of those contigs to the HiPV genome was revealed by BLASTN searches against the viral genomes database. Black triangles indicate protease cleavage sites. VPg, virus protein, genome-linked. (b) Detection of HiPV viral RNA in *L. striatellus* by strand-specific RT-PCR. Each lane represents the RNA sample from a single insect. -RT, no reverse transcription; (+) gRNA and (−) gRNA, PCR amplification from the positive- and negative-strand viral RNA genome, respectively. (c) Quantitative RT-PCR analysis of HiPV RNA accumulation relative to levels of 18S rRNA in insects infected with RBSDV (SI_RB) or RSV (SI_RSV), doubly infected with RBSDV and RSV (DI), or infected with neither virus (VF). Different letters above the bars indicate significant differences (*P*<0.05).
assembled HiPV genome was named HiPV-Js and submitted to NCBI with accession number KC952671. The nucleotide sequence identity between HiPV-Js and the two previously reported HiPV isolates (Guy et al., 1992; Nakashima et al., 2006) was 98.5–98.9%, with amino acid identities >99% in the replicate and 98–98.5% in the capsid protein precursor (analysed using DNAMAN 6).

To determine the abundance of HiPV sRNAs in each of the four libraries, sRNAs (18–30 nt) that perfectly matched the HiPV-Js genome sequence were searched using Bowtie software (http://bowtie-bio.sourceforge.net) with default parameters and the viral sRNA read numbers were normalized with the total reads of the corresponding libraries (scaled to ‘reads per million’). Highly abundant HiPV sRNAs were identified in all libraries and accounted for roughly 7–10% of total sRNAs (Fig. 2a). These sequences were much more abundant than those of RBSDV or RSV identified previously in these libraries (around 1.5 and 0.5%, respectively) (Li et al., 2013). HiPV sRNAs were slightly more abundant in the SI_RBV library than in the VF library, and there were fewer in the SI_RSV and DI libraries (Fig. 2a). The numbers of unique HiPV sRNAs (not identical in sequence to any other sRNAs) were similar in the VF and SI_RBV libraries and also similar, although lower, in the SI_RSV and DI libraries (Fig. 2b). Thus, it seems that there are different effects on the abundance of HiPV sRNAs from infection by RBSDV and RSV. In all libraries, the vast majority of HiPV sRNAs were derived from the positive-strand genome (around 93–95% of total viral sRNAs) (Fig. 2a). Surprisingly, HiPV sRNAs were relatively abundant throughout the size range of 18–30 nt, with a length of 22 nt being slightly more abundant (Fig. 2c for the VF library; data not shown for the SI_RBV, SI_RSV and DI libraries). When the proportions of sense and antisense HiPV sRNAs in each size group were analysed separately, the antisense viral sRNAs were clearly biased towards 21 and 22 nt, which together accounted for more than half of the total HiPV antisense sRNAs (Fig. 2c, d). Thus, the HiPV sense and antisense sRNAs were substantially different in their size distribution patterns. HiPV sense sRNAs from each library were not evenly distributed along the viral genome, with several strong sRNA hotspots, particularly in the 5′-terminal, half of the genome covering the 5′ UTR and replicase-coding region. In contrast, no prominent sRNA hotspots in the viral genome were identified for the antisense sRNAs (Fig. 2e for the VF library; data not shown for the SI_RBV, SI_RSV and DI libraries).

To characterize the HiPV sRNAs further, we performed sequence logo analysis, which calculates the relative frequency of each nucleotide in the sequence through the alignment of multiple sequences (Crooks et al., 2004; http://weblogo.berkeley.edu/logo.cgi). Each size group of HiPV sRNAs was analysed separately to determine the sequence conservations within sRNAs, and the 4 nt proximal to the 5′ and 3′ ends of sRNA in the viral genome sequence were included in the analysis to investigate the possibility that the viral RNA genome is cleaved preferentially at a particular sequence. As presented in Fig. 3, the overall height of the stack indicates the sequence conservation at that position, whilst the height of characters within the stack indicates the relative frequency of nucleotide at that position. No particular nucleotide conservation or motif was found within the sequences of HiPV sense sRNAs. The 5′-terminal nucleotide of HiPV sense sRNAs was most frequently A in the majority (~84%) of size groups, whilst some size groups (24 and 28 nt) most frequently began with U. The nucleotide preceding the 5′-end cleavage site (position −1) was most frequently A and then U (Fig. 3). Strikingly, the 3′ end of HiPV sense sRNAs was most frequently U in all size groups, with the following 4 nt after the cleavage site (positions +1 to +4) most frequently U or to a lesser extent A (Fig. 3 for the VF library; data not shown for the SI_RBV, SI_RSV and DI libraries), showing that their 3′ ends were cleaved preferentially at UA-rich sequences. Taken together, both the 5′ and 3′ ends of HiPV sense sRNAs were most frequently cleaved between dinucleotides UU, UA or AA. In contrast, such a pattern was not observed for HiPV antisense sRNA sequences (Fig. S1) or for those of RBSDV and RSV sRNAs (Fig. S2), showing that this preferential cleavage is a distinct characteristic of HiPV sense sRNAs.

In insects, the RNAi pathway generates 21–23 nt viral siRNA through the activities of Dicer-2 (Sabin et al., 2010). The wide range of size distribution of HiPV sRNAs suggests that another pathway may be involved in their generation in L. striatellus. Some reports have demonstrated that, in Drosophila and mosquitoes, viral 24–30 nt sRNAs are produced by the Dicer-independent PIWI-interacting RNA (piRNA) pathway (Léger et al., 2013; Morazzani et al., 2012; Vodovar et al., 2012; Wu et al., 2010). Our analyses provide some evidence that HiPV sRNAs do not possess the characteristics of piRNAs. First, there was no HiPV sRNA peak at 27 or 28 nt (Fig. 2c) as commonly observed for piRNAs. More importantly, HiPV sRNAs did not possess the typical characteristics of piRNA sequences in which there is a strong U bias at the 5′ terminus or enrichment of A at nt 10 (Fig. 3) (Siomi et al., 2011). Because the RBSDV and RSV sRNAs identified from the same libraries do not have such a wide size distribution (Li et al., 2013), it seems unlikely that the size distribution of HiPV sRNAs is due merely to random degradation of viral genomic RNAs occurring during the isolation of total RNAs from insects. A recent study on sRNAs derived from another crivipavirus, Homalodisca coagulata virus-1 (HoCV-1) in the glassy-winged sharpshooter, Homalodisca vitripennis, also showed that the majority of HoCV-1 sRNAs were sense strand, but a wide range of size distribution was not reported (Nandety et al., 2013). It should be pointed out that, although HiPV antisense sRNAs accounted for only a small portion of total HiPV sRNAs, they were predominantly 21 and 22 nt (Fig. 2c, d), which is characteristic of the products of Dicer processing. This suggests that, to some extent, HiPV sRNAs were generated by RNAi pathways, but that the majority of sense sRNAs were
Fig. 2. Profiles of HiPV sRNAs. (a) Abundance of HiPV sRNA in the four sRNA libraries derived from L. striatellus infected with RBSDV (SI_RB) or RSV (SI_RSV), doubly infected with RBSDV and RSV (DI), or infected with neither of these viruses (VF). (b) Numbers of unique HiPV sRNAs derived from the four libraries. (c) Size distribution of HiPV sRNAs derived from the VF library. (d) Size distribution of HiPV antisense sRNAs. ‘–’ and ‘+’, sRNAs derived, respectively, from the complementary (negative) or positive genomic strands. (e) Distribution of HiPV sRNAs along the HiPV genome. Schematic representations of the HiPV genome organization are presented below the map. Colour coding indicates viral sRNAs derived, respectively, from the positive (+) and negative genomic strands (–). All reads in this analysis were redundant and normalized, except in (b).
Fig. 3. Sequence logo analyses of HiPV sense small RNAs. Each size group of sRNA was separately analysed and the 4 nt proximal to the 5’ and 3’ ends of the sRNAs in the viral genome sequence were included in the analyses.
produced through a different pathway, most probably by cleavage of viral sense genome RNAs by an unknown host RNase. One example of cellular RNase that is directly involved in antiviral defence is human RNase L (RNase L). RNase L is a latent endoribonuclease that is activated by the IFN-induced antiviral 2-5′-oligo-adenylate synthetase pathway (Player & Torrence, 1998). RNase L preferentially cleaves viral RNAs after single-stranded UA and UU dinucleotides (Floyd-Smith et al., 1981; Wreschner et al., 1981). Biochemical analysis is necessary to identify the cellular RNase that is responsible for HiPV RNA degradation. It is also important to further examine whether the HiPV sRNAs have a functional role in antiviral defence or whether they are simply the by-products of viral RNA degradation.

Acknowledgements

The authors are grateful to Dr Mike Adams for critically reading the manuscript. This study was financially supported by the State Basic Research Program of China (2010CB126203), the International Science & Technology Cooperation Program of China (2012DFA30900), the National Natural Science Foundation of China (31071660), Zhejiang Provincial Natural Science Foundation of China (Q12C140016), the Open Foundation from Top Key Discipline of Modern Agricultural Biotechnology and Biological Control of Crop Diseases in Zhejiang Provincial Colleges (2011KFJ004).

References


