Omonrice 15: 108-123 (2007)

EVIDENCE FOR TARGETING COMMON siRNA HOTSPOTS AND GC PREFERENCE BY PLANT DICER-LIKE PROTEINS

Thien Ho[1,2], Hui Wang[1*], Denise Pallett[1] and Tamas Dalmay[3]

[1] NERC/Centre for Ecology and Hydrology (CEH) Oxford, Mansfield Road, Oxford OX1 3SR, UK
[2] Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK
[3] School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK
*CORRESPONDING AUTHOR: Hui Wang, Email: huw@ceh.ac.uk

ABSTRACT

Small interfering (si)RNAs isolated from Brassica juncea leaves infected by Turnip mosaic virus (TuMV) were characterized by cloning and sequencing. The TuMV siRNA population was dominated by 21 and 22-nt long species originated mainly from the same siRNA hotspots, indicating operational similarity between the plant Dicer-like (DCL) enzymes. Robust GC bias was observed for TuMV siRNAs versus the virus genome, indicating that DCL was more likely to target GC-rich regions. Furthermore, dicot micro- miRNAs displayed higher GC% than their DCL1 substrate RNAs, implicating that the GC bias may be ancient, therefore may be important for the RNAi technology.

Keywords: siRNA, miRNA, compositional bias, *Turnip mosaic virus*, *Brassica juncea*, Dicer

INTRODUCTION

Plants have evolved a divergence of Dicer-like (DCL) RNase-III type enzymes that cleave double-stranded dsRNA or single-stranded ssRNA with ds features (Brodersen, *et al.,* 2006*;* Margis *et al.,* 2006). Virus infections trigger plant post-transcriptional gene silencing (PTGS) against viral RNAs, resulting in a virus specific siRNA population in the cytoplasm of plant cells (Baulcombe, 2004). These viral siRNAs are not evenly distributed throughout the virus genomes but are predominately generated in hotspots that may be selected by DCL on ssRNA substrates where fold back structures are formed (Molnar *et al.*, 2005; Moissiard and Voinnet, 2006). Little is known whether DCL may also generate siRNA hotspots when processing dsRNA substrates. In *Arabidopsis*, all DCLs have been recorded as processing viral derived RNAs in coordinated hierarchical actions (Xie *et al.,* 2004; Blevins *et al.,* 2006; Deleris *et al.,*2006; Fusaro *et al.,* 2006; Moissiard and Voinnet, 2006), and DCL2 dependent 22-nt siRNAs were produced redundantly to DCL4 dependent 21-nt siRNAs that mediate anti-viral silencing activities (Blevins *et al.,*2006; Deleris *et al.,*2006; Fusaro *et al.,* 2006). However, little information is available regarding the precise location of different DCL activities on the target viral RNA. This is important because homologous targeting supports the redundancy scenario whereas heterologous targeting suggests variation in function and/or efficiency.

Small RNA cloning and sequencing provides exact siRNA location and homology to the plant virus genome (Ebhardt *et al.,* 2005; Molnar *et al.,* 2005, Ho *et al.,* 2006). However, it is labour intensive and may be affected by 3'-terminal methylations of small RNAs (Ebhardt *et al.,* 2005; Yu *et al.,* 2005). Presence of viral gene silencing suppressors increases the proportion of unmethylated miRNAs in *Arabidopsis* (Yu *et al.,* 2006). The presence of HC-Pro (helper component proteinase) of potyviruses resulted in unmodified 21 and 22-nt viral siRNA populations while other small RNA species were methylated at 3'-termini (Ebhardt *et al.,* 2005). These observations indicate that potyvirus infections provide ideal systems to investigate the viral 21 and 22-nt populations by using the cloning and sequencing approach. Infection with *Turnip mosaic virus* (TuMV, genus *Potyvirus*, single-stranded sense RNA genome of ~9800-nt) results in the production of TuMV siRNAs dominated by 21-nt species in its natural *Brassica* host (Ho *et al.,* 2006), and in *Arabidopsis* (Xie *et al.,* 2004). Here we observed common

siRNA hotspots for the 21 and 22-nt species, from TuMV infected *Brassica juncea cv.* Tendergreen leaves, providing clear evidence that DCLs processed the TuMV RNA at the same siRNA hotspots.

There is a lot of information about small RNA strand asymmetry based on the operational characteristics of Argonaute protein (Hutvagner, 2005). However, little is known about Dicer dependent operational bias. Based on the Watson-Crick base pairing rule, GC content is a determinant of the stability of RNA secondary structures, therefore it may have a functional influence on small RNA biology, particularly Dicer mediated cleavage of dsRNA substrates. In this investigation, we compared the GC% of the TuMV siRNA population to the TuMV genome and detected a robust bias in GC% of the siRNAs versus the viral genome, as well as that of hotspot siRNAs versus non-hotspot ones, providing the first evidence that plant DCLs operate preferably on GC-rich regions.

In contrast to viral siRNAs produced in response to viral infections, micro-(mi)RNAs are encoded by the host, processed by Dicer1 (DCL1), and play important post-transcriptional regulatory roles by targeting mRNAs in both animal and plant kingdoms (Bartel, 2004). We analyzed approximately 4,300 miRNAs available in the miRBase (http://microrna.sanger.ac.uk) and found that there was a robust GC bias in miRNAs in dicot species when the GC% of miRNAs were compared to the GC% of the miRNA precursor sequences with predicted stem-loop structures. This supports the hypothesis that the operational GC bias of DCL in dicots is ancient. On the contrary, miRNAs of all 24 vertebrate species displayed a higher AU% than their precursors, indicating a remarkable divergence between plant and animal RNAi machineries.

MATERIALS AND METHODS

Small RNA cloning and sequencing from TuMV infected plants

Mustard seedlings (*Brassica juncea* cv. tender green) were mechanically inoculated with a pathotype-1 isolate of TuMV (GBR98, isolate and sequence provided by Dr. John Walsh, HRI Warwick, UK) 2 weeks after sowing, and were tested by ELISA to confirm systemic TuMV infections 3 weeks after inoculation (Ho *et al.,* 2006). Young leaves with viral symptoms were sampled and pooled for small RNA cloning and sequencing as previously described (Ho *et al.,* 2006). Leaves of mock- (water) inoculated plants maintained under the same conditions were harvested and processed as negative controls. Sequences of all miRNAs (by January 2007) and their precursor sequences with stem-loop structures were downloaded from the miRBase (http://microrna.sanger.ac.uk).

Data processing and analyses

Small RNA sequences were excised from the adaptor sequences, compared to the virus genome sequence, and had their polarity determined (Ho *et al.,* 2006). Only siRNAs 100% identical to the virus genome were used for further analyses. The 5' end positions of the sense and 3' end positions of the antisense TuMV siRNAs were plotted along the virus genome (MiniTab®14) and siRNA clusters containing more than 10 siRNAs were deemed as siRNA hotspots. GC contents of individual sequences were calculated by FastPCR program (version 3.5.57) (www.biocenter.helsinki.fi/bi/programs/fastpcr.htm). Mean and standard error (SE) were calculated and two-tailed homoscedastic t-tests (Microsoft Excel) were performed to compare the GC% among the small RNAs.

RESULTS AND DISCUSSION

TuMV siRNAs were dominated by 21 and 22-nt species originated from the same hotspots

From leaves of the mock-inoculated plant, 637 small RNA sequences (15-29 nt long, dominated by 21-nt species, Fig. 1A) were obtained, and none resembled the TuMV sequence. Of the 842 small RNA sequences obtained from the TuMV infected leaves, dominated by 21 and 22-nt species (Fig. 1B), 595 were of viral origin with both sense (60%) and antisense (40%) polarity (Fig. 2A). These results

were consistent with observations made from tobacco expressing HC-Pro (Ebhardt *et al.,* 2005), *Arabidopsis* infected with a GFP-tagged-TuMV (TuMV-GFP) (Xie *et al.,* 2004), and the previous report of a significant level of PTGS of the TuMV dsRNA intermediates (Ho *et al.,* 2006). Viral dsRNA intermediates were also implicated as substrates of Dicer-2 in *Drosophila* (Wang *et al.,* 2006). Predictions of RNA secondary structure also supported the view of PTGS targeting of TuMV dsRNA because stem-loop structures were not evident at any of the siRNA hotspot regions in the TuMV genome (Supplementary Fig. 1). The sequenced TuMV siRNA population had a total length of 12,794-nt, far in excess of the virus genome (9,798-nt). When plotted using a window of 100-nt, this siRNA population provided an almost complete coverage of the whole virus genome (Fig. 2B).

The TuMV siRNA population was dominated by species of 21-nt (59.9%), 22-nt (29.1%) and 24-nt (5.0%) in length. The remaining 6% of TuMV siRNAs had lengths of between 16-29 nt (Fig. 2A). When each group of TuMV siRNAs was mapped along the virus genome using a small window of 10-nt, similarities between the profiles of the 21-nt population and 22-nt population were clearly visible for both minus and plus strands (Fig. 2C). Alignments of siRNA sequences at hotspots further revealed that the 21 and 22-nt species were largely overlapping and with site shifts (Supplementary Text 1), indicating that the hotspots were genuine but not generated by cloning procedures (*i.e.*, PCR amplification). The ratio of 21/22-nt at the siRNA hotspots was 60.1% to 28.2% of total (n=219), similar to the overall ratio (n=595) described above, although more 22-nt than 21-nt species were detected in 2/9 hotspots (Hotspot-7, and -9, Supplementary Text 1).

The siRNA size distribution shown in Fig. 2A was comparable with previous observations (in *Arabidopsis*) that the 21-nt long siRNA was the predominant anti-viral silencing component (Blevins *et al.,* 2006; Deleris *et al.,* 2006; Fusaro *et al.,* 2006). Fig. 2C clearly demonstrated that *Brassica* DCLs produced 21-nt long TuMV siRNA in a similar manner to the production of 22-nt TuMV siRNA. This evidence strongly supports previous reports that the *Arabidopsis* DCL2 derived 22-nt species is redundant to the 21-nt species produced by DCL4 (Blevins *et al.,* 2006; Deleris *et al.,* 2006; Fusaro *et al.,* 2006). For the same RNA substrate, if DCLs targeted different locations, it could be suggested that different DCLs have alternative functions and/or different silencing efficiencies. Because abundant 22-nt species were detected in 2/9 TuMV siRNA hotspots, it may suggest that DCL2 could act as the predominant silencing force in some sites despite its overall redundant role to DCL4.

TuMV siRNAs displayed a GC bias

The TuMV siRNA hotspots were strand specific (Fig. 2C), demonstrating Argonaute protein mediated asymmetry (Hutvagner, 2005) in the plant, although the asymmetry of the TuMV siRNAs was apparently not governed by the same rules as animal siRNA production (Ho unpublished data). Analysis of compositional bias of the endogenous small RNAs from the Mock inoculated *B. juncea* sample showed a GC content of 62.3±0.42% (mean±S.E., n=637), considerably higher than the GC%=48.84 of *B. juncea* EST sequences (Codon Usage Database, http://www.kazusa.or.jp/codon/index.html), although a genome based comparison would be more appropriate. To determine whether or not *B. juncea* DCLs may mediate a compositional bias that can not be generated by the strand asymmetry, we calculated the GC content of the sequenced TuMV siRNA populations. The TuMV siRNAs displayed a GC content of 53.6±0.3% (n=595, mean±S.E.) higher than that of the TuMV genome (45.7%). The 9798-nt TuMV genome was then divided into 10 fragments of equal length and the GC% of the viral fragments was compared to the GC% of the siRNAs that had originated from them. The TuMV siRNAs had a higher GC% than the viral genome (P<0.0001, t-tests) (Fig. 3A). A similar story was observed after further dividing the TuMV genome into 98 fragments of 100-nt (Supplementary Fig. 2). These analyses demonstrated that the GC bias was robust and generated by the plant PTGS machinery throughout the whole virus genome in a non-discriminative manner, i.e. the GC bias was due to plant PTGS functions rather than specific viral features. The GC bias was apparent in all TuMV siRNA species of different length and there was no

statistical difference (P>0.1) among the 21-nt (n=357, GC%=53.26±0.33), 22-nt (n=171, GC%=54.07±0.50), 24-nt (n=28, GC%=53.88±0.94), and the other (n=36, GC%=55.00±1.09) species (Fig. 4B). This indicated that the GC bias was not due to any particular DCL activity but was a generic feature of the *Brassica* DCLs. Furthermore, the GC% of the hotspot siRNA population (n=219, GC%=55.71±0.34, Supplementary Text 1) was significantly higher (P<0.0001, tTest) than that of non-hotspot siRNAs (n=376, GC%=52.42±0.34) (Fig. 4C), strongly supporting the theory that GC rich regions of TuMV RNA substrate were preferred by the *Brassica* DCLs. However, GC% did not correlate to the frequency of a siRNA (Supplementary Fig. 2, regression data not shown), indicating that GC% was not the sole determinant of DCL action. It also indicated that the GC bias was not due to the siRNA stability that might be affected by GC%.

Dicot miRNAs also displayed a GC bias

DCL1 is responsible for miRNA production in which mature miRNAs are excised from stem-loop precursor RNA substrates originating from miRNA gene transcripts (Bartel, 2004). Among the 863 plant miRNAs analysed, all (4/4) dicot species had higher (*A. thaliana*, P<0.0001; *Glycine max*, P<0.05; *Medicago truncatula*, P<0.0001; and *Populus trichocarpa*, P<0.0001) GC% in the mature miRNAs than their stem-loop precursor sequences (Fig. 4), specifically supporting that dicot DCL1 have an operational preference for GC rich regions. However, GC bias was not a generic feature for monocot mature miRNAs compared to their stem-loop precursors (Fig. 4). The GC% of monocot miRNAs (n=4, GC%=52.23±1.61) was similar to that of dicot (n=4, GC%=50.32±0.26, P=0.29, t-tests) (Fig. 4). Interestingly, the GC% of miRNA stem-loop precursors in monocots (n=4, GC%=50.80±2.75) was approximately 22% higher than that of dicots (n=4, GC%=41.60±1.60, P<0.05, t-tests) (Fig. 4). It is tempting to speculate that such operational divergence of DCL1 may be a result of DCL diversifications which occurred approximately 200 million years ago between dicots and monocots (Margis *et al.,* 2006). Additionally, because monocots have evolved high GC usage at their 3rd codon positions (GC_3) (Wong *et al.,* 2002; Zhou *et al.,* 2005), it may be possible that the same evolutionary pressure on GC_3 also applied to the miRNA genes. Indeed, it is remarkable that the GC% of miRNA stem-loop precursors in monocots was approximately 22% higher than that of dicots (Fig. 4), considering the small size of these RNA molecules. One of the possible mechanisms that relates miRNA genes to protein evolution is that miRNA genes may originate by inverted duplication of target protein encoding sequences (Allen *et al.,* 2004). As non-conserved miRNA genes may arise from the recent evolutionary past while conserved miRNA genes are considered to be ancient (Jones-Rhoades *et al.,* 2006; Rajagopalan *et al.,* 2006; Fahlgren *et al.*, 2007) we further compared GC% of conserved miRNAs (among *Arabidopsis*, *Populus*, and *Oryza* (Jones-Rhoades *et al.,* 2006) and non-conserved miRNAs to the GC% of their precursors. In all comparisons, mature miRNAs had higher GC% than their precursors (P<0.01, t-tests) (Supplementary Fig. 3). In all of these 3 plant species, GC% of conserved miRNAs was higher than that of non-conserved miRNAs (P<0.0001, t-tests) (Supplementary Fig. 3), strongly supporting that the GC bias is an ancient feature.

It may be logical that DCLs preferably operate on GC rich regions of the RNA substrates because these regions provide a more stable, double-stranded structure, than AU rich regions. Interestingly, animal miRNAs had different compositional bias to the plant miRNAs. Vertebrate species displayed a mild but robust AU bias in the mature miRNAs (n=24, GC%=45.04±0.46) when compared to their stem-loop precursors (n=24, GC%=47.70±0.37, P<0.0001, t-tests) (Fig. 4). Most (6/8) of invertebrate species did not show significant compositional bias (Fig. 4). It may be relevant that the animal systems have evolved to target the AU-rich elements in the 3' untranslated regions of short-lived mRNAs (Jing *et al.,* 2005). Nevertheless, the contradiction between plant and animal miRNAs indicates that operational differences exist between plant and animal Dicers. In dicots, because the GC bias is strong for both siRNA and miRNA (Figs. 3 & 4), it should be considered, in addition to the strand asymmetry, in applications of RNAi.

ACKNOWLEDGMENTS

We are grateful to John Walsh (HRI Warwick, UK) for providing the TuMV isolate and sequence, and to Delia McCall for maintaining plants in glasshouse. This work was supported by the Vietnamese Studentship to TH (Ministry of Education and Training, Decision No 322/QD-TTg), and NERC (UK) grants to TD (NER/A/S/2003/00547) and HW (NER/A/S/2003/00548).

REFERENCES

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W. and Carrington, J.C. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat Genet 36, 1282-90.

Bartel, D.P. (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell 116, 281-297.

Baulcombe, D. (2004) RNA silencing in plants. Nature 431, 356-63.

Blevins, T. *et al*. (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34, 6233-46.

Brodersen, P. and Voinnet, O. (2006) The diversity of RNA silencing pathways in plants. Trends Genet 22, 268-80.

Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K.D., Carrington, J.C. and Voinnet, O. (2006) Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. Science 313, 68-71.

Ebhardt, H.A., Thi, E.P., Wang, M.B. and Unrau, P.J. (2005) Extensive 3' modification of plant small RNAs is modulated by helper component-proteinase expression. Proc Natl Acad Sci USA 102, 13398-403.

Fahlgren, N. *et al*. (2007) High-Throughput Sequencing of Arabidopsis microRNAs: Evidence for Frequent Birth and Death of MIRNA Genes. PLoS ONE 2, e219.

Fusaro, A.F. *et al*. (2006) RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. EMBO Rep 7, 1168-75.

Ho, T., Pallett, D., Rusholme, R., Dalmay, T. and Wang, H. (2006) A simplified method for cloning of short interfering RNAs from Brassica juncea infected with Turnip mosaic potyvirus and Turnip crinkle carmovirus. J Virol Meth 136, 217-223.

Hutvagner, G. (2005) Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. FEBS Lett 579, 5850-7.

Jing, Q. *et al*. (2005) Involvement of microRNA in AU-rich element-mediated mRNA instability. Cell 120, 623-34.

Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. (2006) MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol 57, 19-53.

Margis, R., Fusaro, A.F., Smith, N.A., Curtin, S.J., Watson, J.M., Finnegan, E.J. and Waterhouse, P.M. (2006) The evolution and diversification of Dicers in plants. FEBS Lett 580, 2442-50.

Moissiard, G. and Voinnet, O. (2006) RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins. Proc Natl Acad Sci U S A 103, 19593-8.

Molnar, A., Csorba, T., Lakatos, L., Varallyay, E., Lacomme, C. and Burgyan, J. (2005) Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs. J Virol 79, 7812-7818.

Rajagopalan, R., Vaucheret, H., Trejo, J. and Bartel, D.P. (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev 20, 3407-25.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E. and Carrington, J.C. (2004) Genetic and Functional Diversification of Small RNA Pathways in Plants. PLoS Biology 2, e104.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R. and Chen, X. (2005) Methylation as a crucial step in plant microRNA biogenesis. Science 307, 932-5.

Yu, B., Chapman, E.J., Yang, Z., Carrington, J.C. and Chen, X. (2006) Transgenically expressed viral RNA silencing suppressors interfere with microRNA methylation in Arabidopsis. FEBS Lett 580, 3117-20.

Wang, X.H., Aliyari, R., Li, W.X., Li, H.W., Kim, K., Carthew, R., Atkinson, P. and Ding, S.W. (2006) RNA interference directs innate immunity against viruses in adult Drosophila. Science 312, 452-4.

Wong, G.K., Wang, J., Tao, L., Tan, J., Zhang, J., Passey, D.A. and Yu, J. (2002) Compositional gradients in Gramineae genes. Genome Res 12, 851-6.

Zhou, H., Wang, H., Huang, L.F., Naylor, M. and Clifford, P. (2005) Heterogeneity in codon usages of sobemovirus genes. Arch Virol 150, 1591-605.

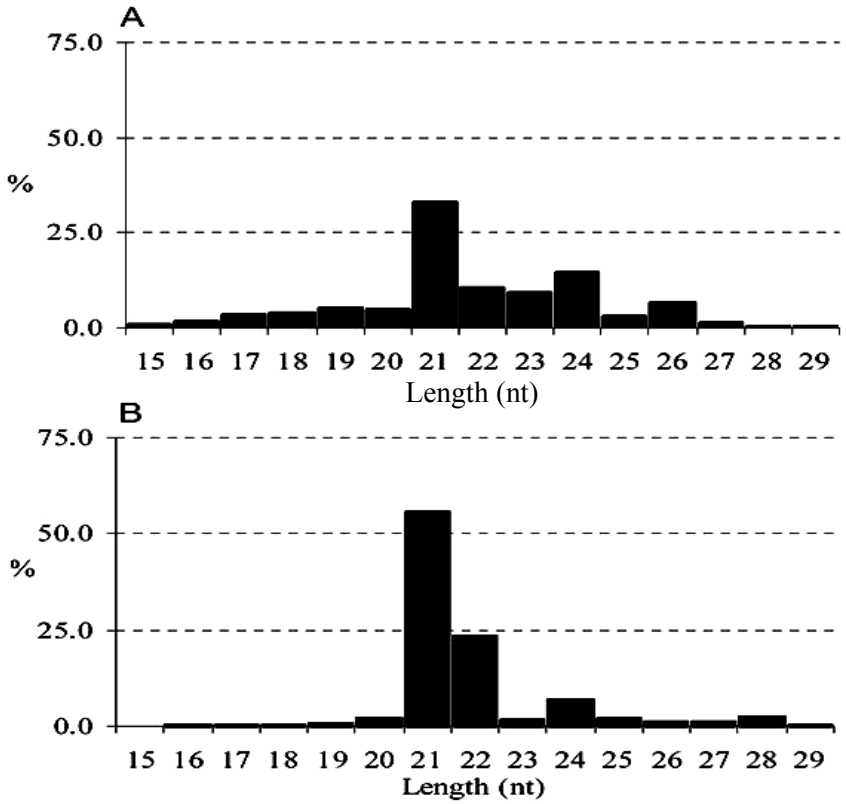


Figure 1. Size distributions of small RNAs

A- Sequenced small RNAs (n=637) from Mock (water inoculated) plants.
B- Sequenced small RNAs (n=842) from TuMV infected plants.

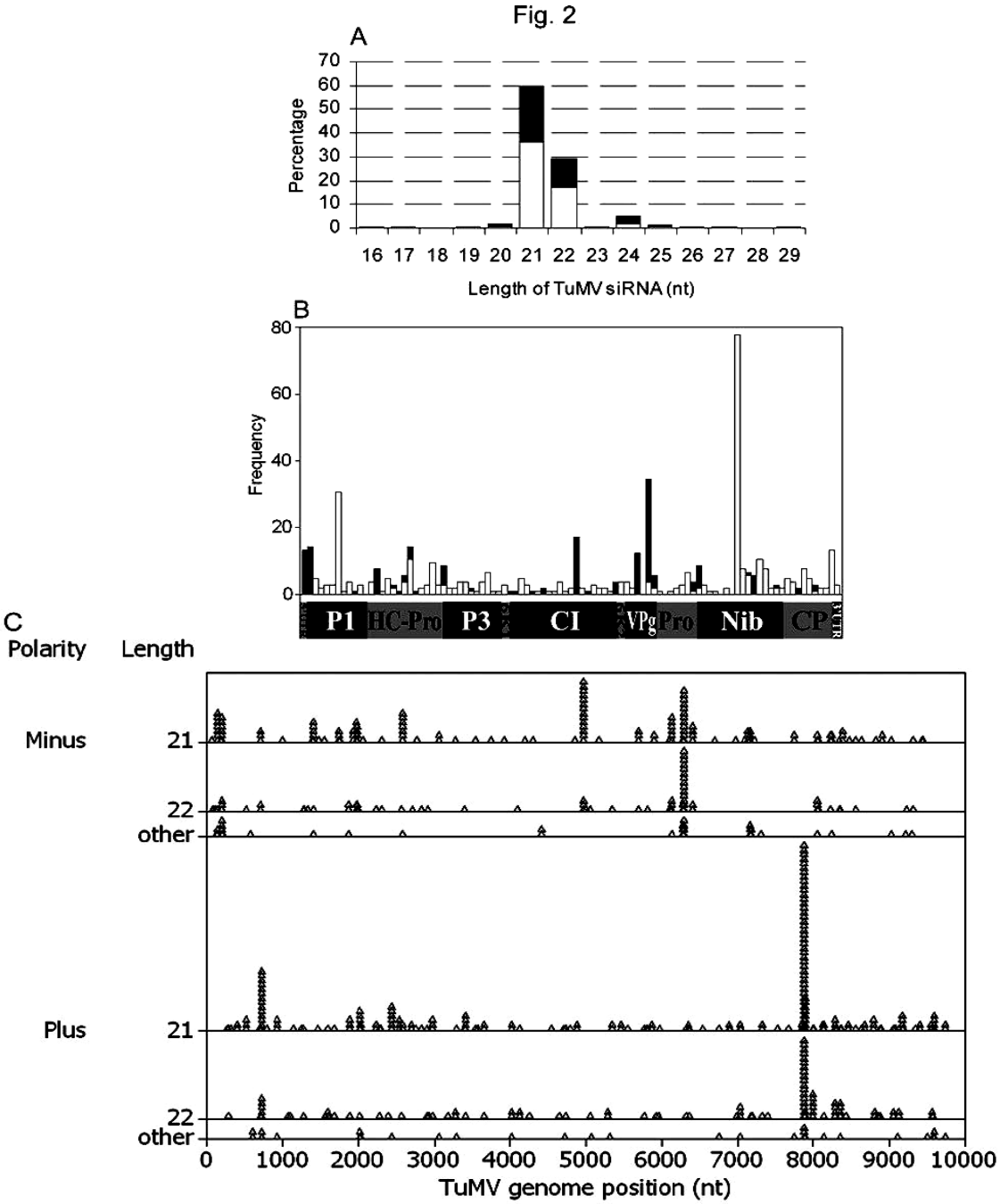


Figure 2. Polarity, size, and location distributions of TuMV siRNAs (n=595)

A. Polarity of TuMV-derived siRNAs in each size class. Y-axis shows the percentage of total TuMV siRNAs with a certain length. Open areas shows the proportion of sense siRNA and filled areas represent that of antisense siRNA.
B. Polarity and genomic distribution of TuMV-derived siRNAs. TuMV siRNAs were plotted (by MiniTab®-14) for each 100-nt window along the virus genome labelled with the viral gene positions. Y-axis shows the number of siRNAs located in each window. Open bars represent sense siRNA, and filled bars represent antisense siRNA.
C. Common hotspots for different size classes of siRNA along the TuMV genome. Each triangle marks the location of a sequenced siRNA, sorted by polarity and length. A small window size of 10-nt was used to identify siRNA clusters, and the alignments of hotspot siRNAs are shown in Supplementary Text 1.

Fig. 3

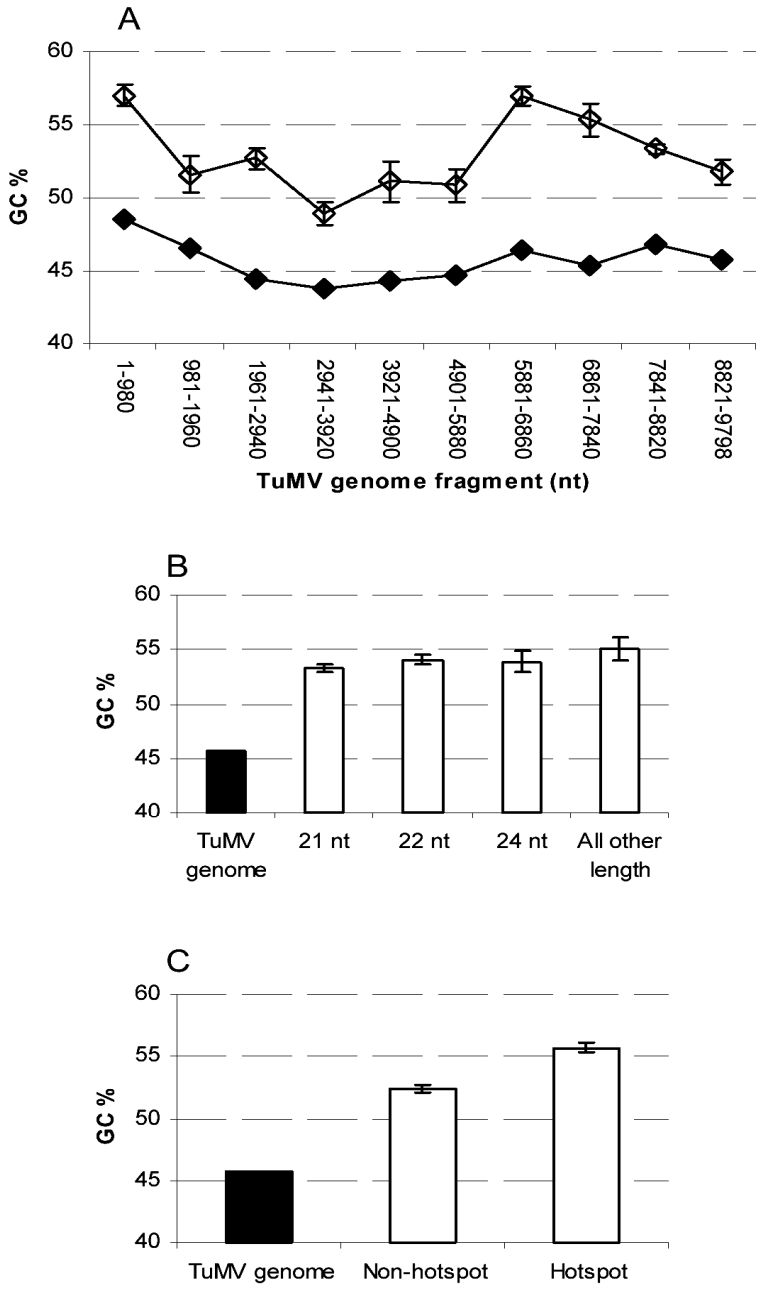


Figure 3. GC bias of TuMV siRNAs

A. GC% of TuMV genomic RNA (filled marks) and siRNA (open marks) was calculated for genome fragments of 980-nt long. The error bars represent SE.
B. Mean GC% of TuMV genomic RNA (filled bar) and siRNA (open bars) of different lengths. The error bars represent SE.
C. GC% of TuMV genomic RNA (filled bar) and siRNA (open bars) from hotspots or non-hotspots. The error bars represent SE.

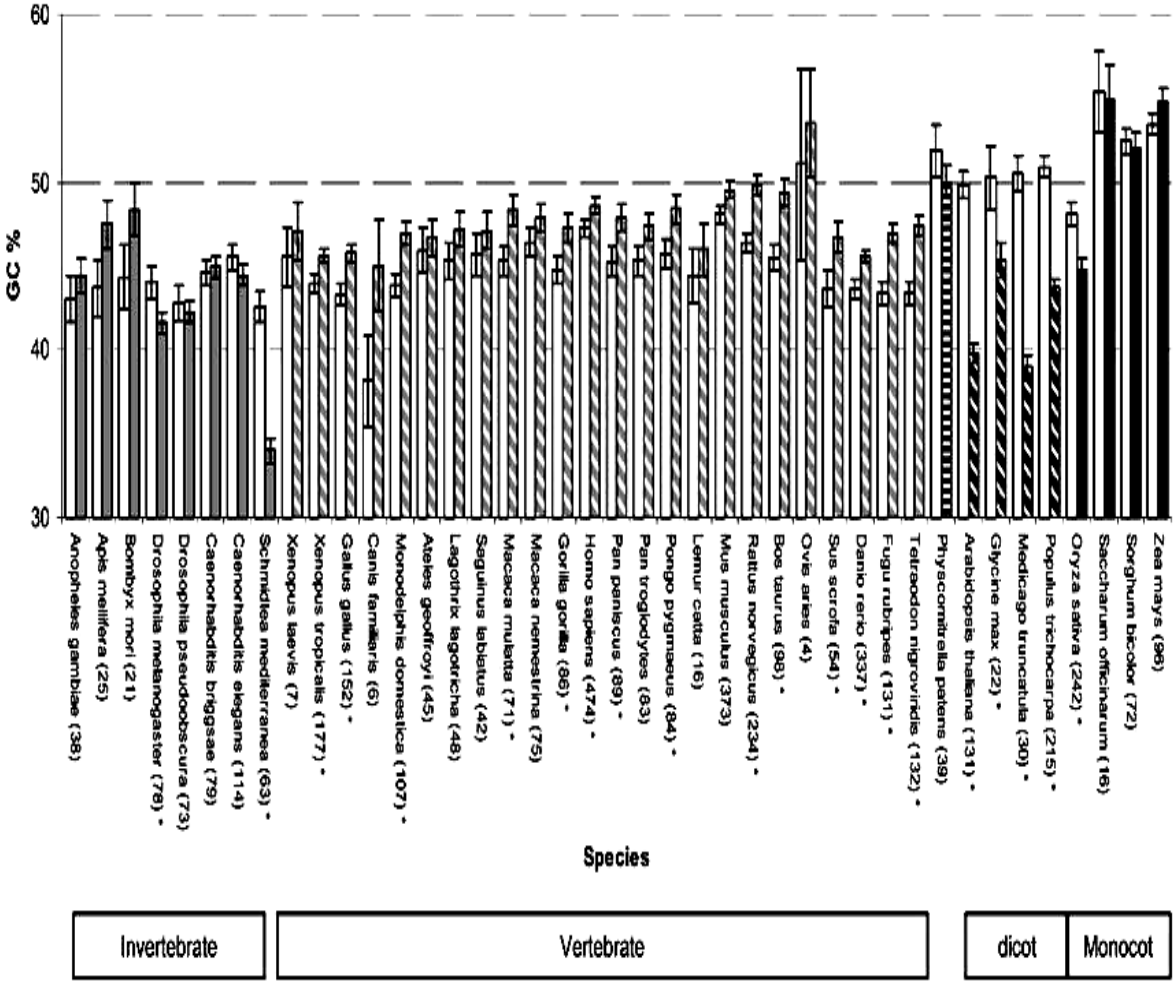
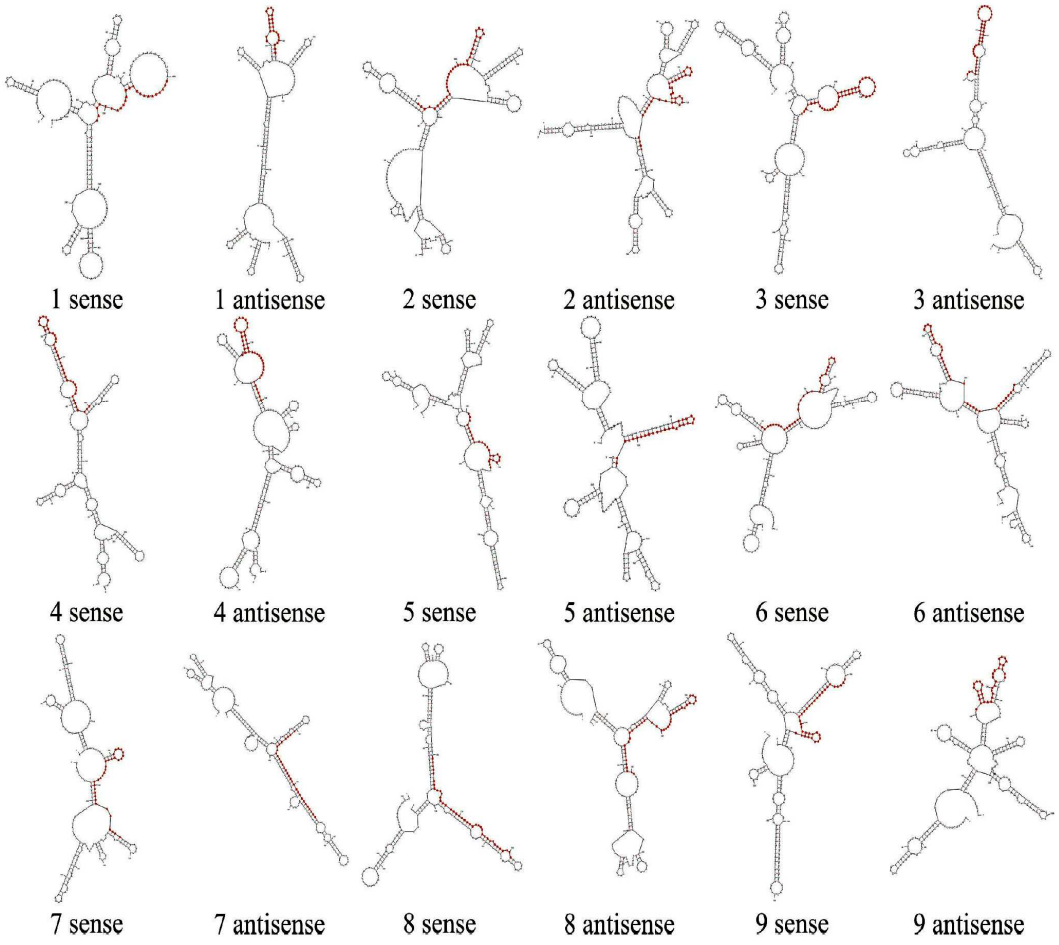


Figure 4. GC contents of miRNAs

GC% (mean±SE) of miRNAs (open bars) and their precusors (filled bars) of invertebrate, vertebrate, dicot, and monocot species is presented. The number of miRNA sequences is shown after a species name. Asterisk labels significant difference (P<0.05) between miRNA and corresponding precursor.

Supplementary materials

Supplementary Fig. 1

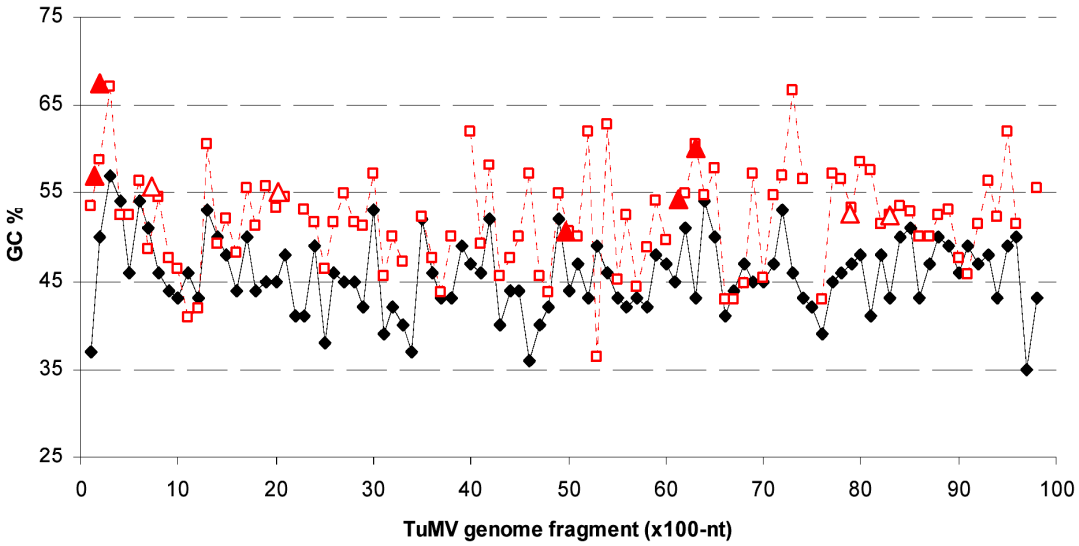


Supplementary Fig. 1: Predicted TuMV ssRNA structures at siRNA hotspots

RNA folding predictions were performed by feeding the TuMV sequences flanking the siRNA hotspots (labeled by red letters, and see sequence details in Supplementary Text 1) into the Mfold program (version 3.2) (http://www.bioinfo.rpi.edu/applications/mfold/) [24,25] with default setting.

[24] Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288, 911-940.
[25] Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31, 3406-3415.

Supplementary Fig. 2

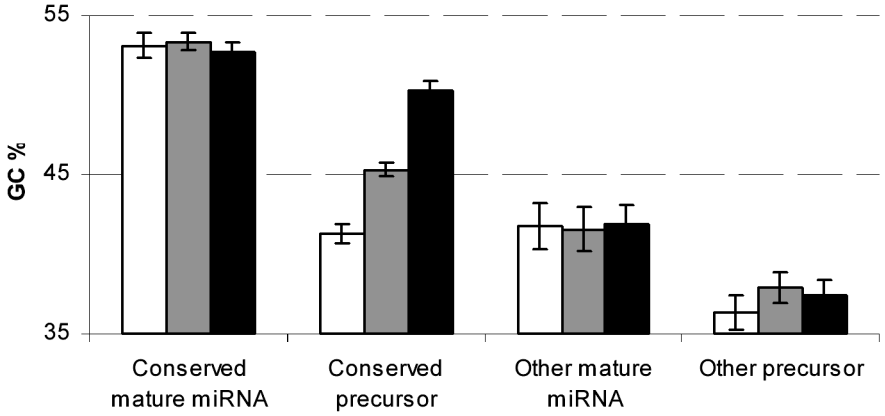


Supplementary Fig. 2: GC bias of TuMV siRNA using 100-nt windows

The TuMV genome was devided into 98 fragments of 100-nt in length. GC% of the TuMV fragments (black sold line labelled by filled diamonds) and the GC% of the siRNAs (red dashed line labelled by unfilled squares) that had originated from them were plotted along the TuMV genome positions. The siRNA line had broken points where no siRNA was detected in the particular TuMV fragments. The GC% of siRNA hotspot clusters (Supplementary Text 1) were plotted for sense (open triangle) and anti-sense (filled triangle) origins accordingly to their locations in the TuMV genome using the midpoint of the clusters. The GC% of siRNAs was significantly higher than that of the TuMV fragments (P<0.0001, t-test)

Supplementary Figure 2. GC bias of TuMV siRNA using 100-nt windows

Supplementary Fig. 3



Supplementary Fig. 3: GC bias of conserved and non-conserved plant miRNAs

GC contents (mean±S.E.) were calculated for conserved miRNAs and their precursor sequences in *Arabidopsis thaliana* (open bars), *Populus trichocarpa* (grey bars), and *Oryza sativa* (black bars) [20]. Comparisons were made to GC% of the other miRNAs from the same plant species (used in Fig.4). Mature miRNAs displayed higher GC% than miRNA precursors in both conserved and non-conserved classes (P<0.01). In all of the three plant species, GC% of the conserved miRNAs were significantly higher than GC% of the other miRNAs available in miRBase (P<0.0001).

[20] Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. (2006) MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol 57, 19-53.

Supplementary Figure 3. GC bias of conserved and non-conserved miRNAs

Supplementary Text 1: Sequence alignments of hotspot siRNAs

Supplementary Text 1: Sequence alignments of hotspot TuMV siRNAs

The hotspot siRNA sequences were aligned by the ClustalX program (ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) and manually checked.

Hotspot 1

CLUSTAL X (1.8) multiple sequence alignment

```
Name      Length  Polarity    GC%   Position  Sequence (3'← 5')
tu202.1  24       Minus            54.17    138       TTCCGGTGAGTGCTGGTTTGTTGG
tu140.7  22       Minus            59.09    138       --CCGGTGAGTGCTGGTTTGTTGG
tu89.2   21       Minus            57.14    138       ---CGGTGAGTGCTGGTTTGTTGG
tu222.3  21       Minus            57.14    138       ---CGGTGAGTGCTGGTTTGTTGG
tu206.3  21       Minus            57.14    138       ---CGGTGAGTGCTGGTTTGTTGG
tu203.3  21       Minus            57.14    138       ---CGGTGAGTGCTGGTTTGTTGG
tu195.2  21       Minus            57.14    138       ---CGGTGAGTGCTGGTTTGTTGG
tu175.1  21       Minus            57.14    138       ---CGGTGAGTGCTGGTTTGTTGG
tu153.10 21       Minus            57.14    138       ---CGGTGAGTGCTGGTTTGTTGG
tu150.5  20       Minus            55.00    139       ---CGGTGAGTGCTGGTTTGTTG-
                                                      ********************
```

Hotspot 2

CLUSTAL X (1.8) multiple sequence alignment

```
Name      Length  Polarity    GC%   Position  Sequence (3'← 5')
tu74.2   21       Minus            71.43    196       ---------------GCGGTGACGGTGGTGGATCGC
tu120.13 21       Minus            61.90    199       ------------ACTGCGGTGACGGTGGTGGAT---
tu93.2   22       Minus            68.18    200       ----------CGACTGCGGTGACGGTGGTGGA----
tu152.1  22       Minus            68.18    200       ----------CGACTGCGGTGACGGTGGTGGA----
tu179.4  21       Minus            71.43    201       ----------CGACTGCGGTGACGGTGGTGG-----
tu115.1  20       Minus            70.00    202       ----------CGACTGCGGTGACGGTGGTG------
tu110.3  21       Minus            71.43    202       ---------GCGACTGCGGTGACGGTGGTG------
tu190.6  21       Minus            71.43    202       ---------GCGACTGCGGTGACGGTGGTG------
tu219.1  26       Minus            65.38    204        -TGAAGTGGCGACTGCGGTGACGGTGG--------
tu71.1   25       Minus            64.00    206        CTGAAGTGGCGACTGCGGTGACGGT----------
tu175.7  24       Minus            66.67    207        CTGAAGTGGCGACTGCGGTGACGG-----------
tu91.9   22       Minus            68.18    207        --GAAGTGGCGACTGCGGTGACGG-----------
tu49.1   21       Minus            66.67    207        ---AAGTGGCGACTGCGGTGACGG-----------
tu194.8  21       Minus            66.67    208        --GAAGTGGCGACTGCGGTGACG------------
tu189.7  20       Minus            60.00    210        -TGAAGTGGCGACTGCGGTGA--------------
                                                             *******
```

Hotspot 3

CLUSTAL X (1.8) multiple sequence alignment

```
Name      Length  polarity    GC%   Position  Sequence (5'→ 3')
tu94.10  21       Plus             52.38    711       CACCGTCAATGAGGAAGAGGA----------------------
tu173.4  22       Plus             54.55    716       -----TCAATGAGGAAGAGGACGGTGC---------------
tu148.1  22       Plus             54.55    716       -----TCAATGAGGAAGAGGACGGTGC---------------
tu55.1   21       Plus             57.14    717       ------CAATGAGGAAGAGGACGGTGC---------------
tu114.10 21       Plus             57.14    717       ------CAATGAGGAAGAGGACGGTGC---------------
tu23.5   21       Plus             52.38    718       -------AATGAGGAAGAGGACGGTGCA--------------
tu192.7  21       Plus             52.38    718       -------AATGAGGAAGAGGACGGTGCA--------------
tu192.6  21       Plus             52.38    718       -------AATGAGGAAGAGGACGGTGCA--------------
tu95.8   21       Plus             57.14    719       --------ATGAGGAAGAGGACGGTGCAC-------------
tu196.2  19       Plus             57.89    720       ---------TGAGGAAGAGGACGGTGCA--------------
tu161.2  19       Plus             57.89    720       ---------TGAGGAAGAGGACGGTGCA--------------
tu159.7  21       Plus             57.14    720       ---------TGAGGAAGAGGACGGTGCACA------------
tu153.3  21       Plus             57.14    723       ------------GGAAGAGGACGGTGCACAGAA---------
tu89.5   21       Plus             57.14    723       ------------GGAAGAGGACGGTGCACAGAA---------
tu161.11 22       Plus             54.55    724       -------------GAAGAGGACGGTGCACAGAACA-------
tu186.12 22       Plus             54.55    724       -------------GAAGAGGACGGTGCACAGAACA-------
tu196.11 22       Plus             54.55    724       -------------GAAGAGGACGGTGCACAGAACA-------
tu125.10 21       Plus             57.14    724       -------------GAAGAGGACGGTGCACAGAAC--------
tu191.7  21       Plus             57.14    724       -------------GAAGAGGACGGTGCACAGAAC--------
tu147.5  21       Plus             52.38    725       --------------AAGAGGACGGTGCACAGAACA-------
tu188.6  21       Plus             52.38    725       --------------AAGAGGACGGTGCACAGAACA-------
tu35.6   21       Plus             52.38    725       --------------AAGAGGACGGTGCACAGAACA-------
tu199.2  21       Plus             57.14    727       ----------------GAGGACGGTGCACAGAACATG-----
tu180.4  21       Plus             57.14    727       ----------------GAGGACGGTGCACAGAACATG-----
tu176.7  21       Plus             57.14    727       ----------------GAGGACGGTGCACAGAACATG-----
tu129.2  21       Plus             57.14    727       ----------------GAGGACGGTGCACAGAACATG-----
tu104.2  21       Plus             57.14    727       ----------------GAGGACGGTGCACAGAACATG-----
tu125.11 21       Plus             57.14    728       -----------------AGGACGGTGCACAGAACATGC----
tu61.1   24       Plus             54.17    729       -----------------GGACGGTGCACAGAACATGCAAGA
                                                              ***
```

Hotspot 4
CLUSTAL X (1.8) multiple sequence alignment

```
Name      Length  polarity    GC%   Position  Sequence (5′→ 3′)
tu127.4  24       Plus             58.33    2014      CGAGTGTGTGAGCAAGTTGCAGGG----------------
tu118.8  25       Plus             52.00    2016      --AGTGTGTGAGCAAGTTGCAGGGTGA--------------
tu220.1  22       Plus             54.55    2018      ----TGTGTGAGCAAGTTGCAGGGTG---------------
tu73.3   21       Plus             57.14    2019      -----GTGTGAGCAAGTTGCAGGGTG---------------
tu68.2   21       Plus             57.14    2019      -----GTGTGAGCAAGTTGCAGGGTG---------------
tu225.8  21       Plus             57.14    2019      -----GTGTGAGCAAGTTGCAGGGTG---------------
tu202.5  21       Plus             57.14    2019      -----GTGTGAGCAAGTTGCAGGGTG---------------
tu186.5  21       Plus             52.38    2020      ------TGTGAGCAAGTTGCAGGGTGA--------------
tu39.5   24       Plus             50.00    2024      ----------AGCAAGTTGCAGGGTGACTTTGTC------
tu144.7  21       Plus             52.38    2031      -----------------TGCAGGGTGACTTTGTCCATG--
tu90.4   21       Plus             57.14    2033      -------------------CAGGGTGACTTTGTCCATGCG
                                                                 *****
```

Hotspot 5
CLUSTAL X (1.8) multiple sequence alignment

```
Name      Length  Polarity    GC%   Position   Sequence (3'← 5′)
tu25.8   21       Minus            47.62    4965       ----TCGAGACGTTGTTGGTGATCA
tu131.10 21       Minus            47.62    4965       ----TCGAGACGTTGTTGGTGATCA
tu226.1  21       Minus            52.38    4966       ---GTCGAGACGTTGTTGGTGATC-
tu81.4   22       Minus            50.00    4966       --TGTCGAGACGTTGTTGGTGATC-
tu66.9   22       Minus            50.00    4966       --TGTCGAGACGTTGTTGGTGATC-
tu57.9   22       Minus            50.00    4966       --TGTCGAGACGTTGTTGGTGATC-
tu137.11 21       Minus            47.62    4968       -TTGTCGAGACGTTGTTGGTGA---
tu194.1  21       Minus            47.62    4968       -TTGTCGAGACGTTGTTGGTGA---
tu57.2   21       Minus            47.62    4968       -TTGTCGAGACGTTGTTGGTGA---
tu146.2  21       Minus            52.38    4969       CTTGTCGAGACGTTGTTGGTG----
tu102.5  21       Minus            52.38    4969       CTTGTCGAGACGTTGTTGGTG----
tu164.4  21       Minus            52.38    4969       CTTGTCGAGACGTTGTTGGTG----
tu181.8  21       Minus            52.38    4969       CTTGTCGAGACGTTGTTGGTG----
tu35.5   21       Minus            52.38    4969       CTTGTCGAGACGTTGTTGGTG----
tu56.3   21       Minus            52.38    4969       CTTGTCGAGACGTTGTTGGTG----
tu69.2   21       Minus            52.38    4969       CTTGTCGAGACGTTGTTGGTG----
tu78.4   21       Minus            52.38    4969       CTTGTCGAGACGTTGTTGGTG----
                                                      ****************
```

Hotspot 6
CLUSTAL X (1.8) multiple sequence alignment

```
Name      Length  polarity    GC%   Position  Sequence (3′← 5′)
tu115.3  22       Minus            54.55    6122      --------------------TGATGTCTGTGAGCGGGTTGTC
tu195.7  21       Minus            61.90    6127      ---------------AGGGTGATGTCTGTGAGCGGG-----
tu111.1  21       Minus            57.14    6129      -------------CAAGGGTGATGTCTGTGAGCG--------
tu18.3   22       Minus            54.55    6130      -----------CACAAGGGTGATGTCTGTGAGC---------
tu118.2  20       Minus            55.00    6130      -------------CAAGGGTGATGTCTGTGAGC---------
tu158.4  22       Minus            54.55    6131      ----------GCACAAGGGTGATGTCTGTGAG----------
tu160.4  21       Minus            52.38    6133      ---------TGCACAAGGGTGATGTCTGTG-----------
tu163.4  21       Minus            52.38    6133      ---------TGCACAAGGGTGATGTCTGTG-----------
tu66.8   22       Minus            50.00    6135      ------TCTTGCACAAGGGTGATGTCTG--------------
tu57.11  21       Minus            52.38    6135      -------CTTGCACAAGGGTGATGTCTG--------------
tu218.4  21       Minus            52.38    6135      -------CTTGCACAAGGGTGATGTCTG--------------
                                                                 *********
```

Hotspot 7
CLUSTAL X (1.8) multiple sequence alignment

```
Name      Length  polarity    GC%   Position   Sequence (3′← 5′)
tu78.10  29       Minus            62.07    6280       --CCTTGAGGGGCATGTGTGGTGTCAGGTCC
tu171.1  25       Minus            64.00    6280       ------GAGGGGCATGTGTGGTGTCAGGTCC
tu223.3  24       Minus            62.50    6280       -------AGGGGCATGTGTGGTGTCAGGTCC
tu96.1   21       Minus            61.90    6281       ----------GGGCATGTGTGGTGTCAGGTC-
tu180.7  22       Minus            63.64    6281       --------GGGGCATGTGTGGTGTCAGGTC-
tu115.8  22       Minus            63.64    6281       --------GGGGCATGTGTGGTGTCAGGTC-
tu74.11  24       Minus            58.33    6282       -----TGAGGGGCATGTGTGGTGTCAGGT--
tu66.7   22       Minus            59.09    6282       -------AGGGGCATGTGTGGTGTCAGGT--
tu157.1  22       Minus            59.09    6282       -------AGGGGCATGTGTGGTGTCAGGT--
tu197.2  21       Minus            61.90    6282       --------GGGGCATGTGTGGTGTCAGGT--
tu169.10 24       Minus            58.33    6283       ----TTGAGGGGCATGTGTGGTGTCAGG---
tu196.6  21       Minus            61.90    6283       -------AGGGGCATGTGTGGTGTCAGG---
tu161.6  21       Minus            61.90    6283       -------AGGGGCATGTGTGGTGTCAGG---
tu96.4   22       Minus            63.64    6283       ------GAGGGGCATGTGTGGTGTCAGG---
tu86.1   22       Minus            63.64    6283       ------GAGGGGCATGTGTGGTGTCAGG---
tu143.3  22       Minus            63.64    6283       ------GAGGGGCATGTGTGGTGTCAGG---
tu140.3  22       Minus            63.64    6283       ------GAGGGGCATGTGTGGTGTCAGG---
tu122.4  22       Minus            63.64    6283       ------GAGGGGCATGTGTGGTGTCAGG---
tu148.3  22       Minus            63.64    6283       ------GAGGGGCATGTGTGGTGTCAGG---
tu33.1   27       Minus            59.26    6284       CACCTTGAGGGGCATGTGTGGTGTCAG----
```

```
tu79.4   22       Minus            59.09    6284       -----TGAGGGGCATGTGTGGTGTCAG----
tu4.6    22       Minus            59.09    6284       -----TGAGGGGCATGTGTGGTGTCAG----
tu193.1  21       Minus            61.90    6284       ------GAGGGGCATGTGTGGTGTCAG----
tu71.8   21       Minus            61.90    6284       ------GAGGGGCATGTGTGGTGTCAG----
tu141.5  21       Minus            57.14    6285       -----TGAGGGGCATGTGTGGTGTCA-----
tu223.1  22       Minus            59.09    6286       ---CTTGAGGGGCATGTGTGGTGTC------
tu64.1   21       Minus            57.14    6286       ----TTGAGGGGCATGTGTGGTGTC------
tu57.5   21       Minus            57.14    6286       ----TTGAGGGGCATGTGTGGTGTC------
tu32.2   21       Minus            57.14    6286       ----TTGAGGGGCATGTGTGGTGTC------
tu213.2  21       Minus            57.14    6286       ----TTGAGGGGCATGTGTGGTGTC------
tu102.1  21       Minus            57.14    6286       ----TTGAGGGGCATGTGTGGTGTC------
tu59.1   24       Minus            58.33    6287       CACCTTGAGGGGCATGTGTGGTGT-------
tu124.7  22       Minus            59.09    6288       -ACCTTGAGGGGCATGTGTGGTG--------
tu124.3  22       Minus            59.09    6288       -ACCTTGAGGGGCATGTGTGGTG--------
tu102.6  22       Minus            59.09    6289       CACCTTGAGGGGCATGTGTGGT---------
tu114.9  21       Minus            61.90    6290       CACCTTGAGGGGCATGTGTGG----------
                                                          ************
```

Hotspot 8

CLUSTAL X (1.8) multiple sequence alignment

```
Name_Length_polarity_GC%_Position_Sequence (5'→ 3')
tu226.5_25__Plus 60.00    7876      GGAGTCGTTGCCAGATGGTTGGGTG----------------
tu202.4_21__Plus 57.14    7879      ---GTCGTTGCCAGATGGTTGGGT-----------------
tu37.1_22__Plus  59.09    7879      ---GTCGTTGCCAGATGGTTGGGTG----------------
tu149.2_22_Plus  59.09    7879      ---GTCGTTGCCAGATGGTTGGGTG----------------
tu115.4_22_Plus  59.09    7879      ---GTCGTTGCCAGATGGTTGGGTG----------------
tu107.1_22_Plus  59.09    7879      ---GTCGTTGCCAGATGGTTGGGTG----------------
tu136.9_21_Plus  57.14    7880      ----TCGTTGCCAGATGGTTGGGTG----------------
tu56.4_22_Plus   54.55    7880      ----TCGTTGCCAGATGGTTGGGTGT---------------
tu5.9_21_Plus    57.14    7881      -----CGTTGCCAGATGGTTGGGTGT---------------
tu185.4_20_Plus  55.00    7882      ------GTTGCCAGATGGTTGGGTGT---------------
tu86.8_22_Plus   54.55    7881      -----CGTTGCCAGATGGTTGGGTGTA--------------
tu219.9_22_Plus  54.55    7881      -----CGTTGCCAGATGGTTGGGTGTA--------------
tu202.2_22_Plus  54.55    7881      -----CGTTGCCAGATGGTTGGGTGTA--------------
tu191.4_22_Plus  54.55    7881      -----CGTTGCCAGATGGTTGGGTGTA--------------
tu186.1_22_Plus  54.55    7881      -----CGTTGCCAGATGGTTGGGTGTA--------------
tu133.1_22_Plus  54.55    7881      -----CGTTGCCAGATGGTTGGGTGTA--------------
tu223.8_25_Plus  52.00    7881      -----CGTTGCCAGATGGTTGGGTGTATTG-----------
tu79.1_21_Plus   52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu70.6_21_Plus   52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu66.2_21_Plus   52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu53.4_21_Plus   52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu52.7_21_Plus   52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu51.3_21_Plus   52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu222.2_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu211.4_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu208.8_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu204.8_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu183.8_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu174.11_21_Plus 52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu168.8_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu151.4_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu147.3_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu136.10_21_Plus 52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu132.6_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu120.9_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu117.5_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu114.4_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu10.4_21_Plus   52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu109.5_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu174.7_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu222.1_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu225.3_21_Plus  52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu4.3_21_Plus    52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu86.3_21_Plus   52.38    7882      ------GTTGCCAGATGGTTGGGTGTA--------------
tu225.2_22_Plus  50.00    7882      ------GTTGCCAGATGGTTGGGTGTAT-------------
tu53.1_21_Plus   47.62    7883      -------TTGCCAGATGGTTGGGTGTAT-------------
tu33.3_21_Plus   47.62    7883      -------TTGCCAGATGGTTGGGTGTAT-------------
tu140.8_22_Plus  50.00    7884      --------TGCCAGATGGTTGGGTGTATTG-----------
tu178.6_22_Plus  50.00    7884      --------TGCCAGATGGTTGGGTGTATTG-----------
tu194.7_22_Plus  50.00    7884      --------TGCCAGATGGTTGGGTGTATTG-----------
tu36.2_22_Plus   50.00    7884      --------TGCCAGATGGTTGGGTGTATTG-----------
tu85.4_22_Plus   50.00    7884      --------TGCCAGATGGTTGGGTGTATTG-----------
tu204.4_21_Plus  52.38    7885      ---------GCCAGATGGTTGGGTGTATTG-----------
tu220.8_21_Plus  52.38    7885      ---------GCCAGATGGTTGGGTGTATTG-----------
tu33.2_21_Plus   52.38    7885      ---------GCCAGATGGTTGGGTGTATTG-----------
tu137.12_17_Plus 58.82    7885      ---------GCCAGATGGTTGGGTGT---------------
tu131.4_21_Plus  52.38    7886      ----------CCAGATGGTTGGGTGTATTGC----------
tu25.2_21_Plus   52.38    7886      ----------CCAGATGGTTGGGTGTATTGC----------
tu10.3_21_Plus   52.38    7887      -----------CAGATGGTTGGGTGTATTGCG---------
```

```
tu226.3_21_Plus  52.38    7887      -----------CAGATGGTTGGGTGTATTGCG---------
tu226.4_21_Plus  52.38    7887      -----------CAGATGGTTGGGTGTATTGCG---------
tu88.2_21_Plus   52.38    7887      -----------CAGATGGTTGGGTGTATTGCG---------
tu94.4_21_Plus   47.62    7888      ------------AGATGGTTGGGTGTATTGCGA--------
tu183.6_22_Plus  50.00    7889      -------------GATGGTTGGGTGTATTGCGATG------
tu80.3_20_Plus   50.00    7891      ---------------TGGTTGGGTGTATTGCGATG------
tu136.4_21_Plus  52.38    7891      ---------------TGGTTGGGTGTATTGCGATGC-----
tu151.9_21_Plus  52.38    7891      ---------------TGGTTGGGTGTATTGCGATGC-----
tu176.4_21_Plus  52.38    7893      -----------------GTTGGGTGTATTGCGATGCTG---
tu187.6_21_Plus  52.38    7893      -----------------GTTGGGTGTATTGCGATGCTG---
tu23.3_21_Plus   52.38    7893      -----------------GTTGGGTGTATTGCGATGCTG---
tu183.5_22_Plus  50.00    7893      -----------------GTTGGGTGTATTGCGATGCTGA--
tu47.1_21_Plus   47.62    7894      ------------------TTGGGTGTATTGCGATGCTGA--
tu146.9_21_Plus  47.62    7894      ------------------TTGGGTGTATTGCGATGCTGA--
tu117.2_21_Plus  47.62    7894      ------------------TTGGGTGTATTGCGATGCTGA--
tu145.2_22_Plus  50.00    7895      ------------------TGGGTGTATTGCGATGCTGATG
                                                     *****
```

Hotspot 9

```
CLUSTAL X (1.8) multiple sequence alignment

Name      Length  polarity    GC%   Position  Sequence (5′→ 3′)
tu140.9  22       Plus             54.55    8284      CAACTTTCGTGAACTGGGCCTG-------------
tu129.3  22       Plus             54.55    8284      CAACTTTCGTGAACTGGGCCTG-------------
tu141.8  22       Plus             50.00    8285      -AACTTTCGTGAACTGGGCCTGA------------
tu225.5  22       Plus             54.55    8287      ---CTTTCGTGAACTGGGCCTGAAG----------
tu86.2   21       Plus             52.38    8287      ---CTTTCGTGAACTGGGCCTGAA-----------
tu147.2  22       Plus             50.00    8289      -----TTCGTGAACTGGGCCTGAAGTA--------
tu134.5  21       Plus             52.38    8291      -------CGTGAACTGGGCCTGAAGTAT-------
tu201.5  21       Plus             52.38    8291      -------CGTGAACTGGGCCTGAAGTAT-------
tu88.5   22       Plus             50.00    8291      -------CGTGAACTGGGCCTGAAGTATA------
tu135.2  22       Plus             54.55    8297      -------------CTGGGCCTGAAGTATACCTTCG
                                                            *********
```

Bằng chứng của việc tấn công vào cùng các điểm nóng siRNA và sự chọn lọc GC% của các protein tương tự DICER

Các phân tử RNA ngắn có chức năng cản trở (small interfering (si) RNA) từ lá của cây *Brassica juncea* bị nhiễm bởi *Turnip mosaic virus* (TuMV) đã được ly trích, đọc mã và phân tích. Hầu hết các phân tử siRNA này có chiều dài 21-22 nucleotide và thường tạo thành các 'điểm nóng siRNA'. Điều này cho thấy rằng có thể có sự tương đồng trong quá trình hoạt động của các enzyme tương-tự-như-DICER (DCL) của thực vật. Vì có sự khác biệt lớn về GC% giữa các phân tử siRNA và bộ gene của virus, do vậy các DCL có thể thường tấn công vào các vùng chứa nhiều GC. Ngoài ra, do các phân tử microRNA có GC% cao hơn các phân tử RNA là sản phẩm của DCL1, cho nên sự khác biệt về GC có thể có nguồn gốc cổ xưa. Sự phát hiện này có khả năng áp dụng quan trọng trong kỹ thuật *cản trở RNA* (RNA interference).