FEBS Letters 581 (2007) 3267–3272

Evidence for targeting common siRNA hotspots and GC preference by plant Dicer-like proteins

Thien Ho[a,b], Hui Wang[a,*], Denise Pallett[a], Tamas Dalmay[c]

[a] *NERC/Centre for Ecology and Hydrology (CEH) Oxford, Mansfield Road, Oxford OX1 3SR, UK*
[b] *Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK*
[c] *School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK*

Received 27 April 2007; revised 1 June 2007; accepted 13 June 2007

Available online 21 June 2007

Edited by Shou-Wei Ding

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-(mi)RNAs 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; Dicer; *Turnip mosaic virus*; *Brassica juncea*

1. Introduction

Plants have evolved a divergence of Dicer-like (DCL) RNase-III type enzymes that cleave double-stranded (ds)RNA or single-stranded (ss)RNA with ds features [1,2]. 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 [3]. 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 [4,5]. 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 [5–9], and DCL2 dependent 22-nt siRNAs were produced redundantly to DCL4 dependent 21-nt siRNAs that mediate anti-viral silencing activities [7–9]. 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.

*Corresponding author. Fax: +44 01865 281696.
E-mail addresses: thien.ho@pembroke.oxford.ac.uk (T. Ho), huw@ ceh.ac.uk (H. Wang), dwpa@ceh.ac.uk (D. Pallett), t.dalmay@uea. ac.uk (T. Dalmay).

Small RNA cloning and sequencing provides exact siRNA location and homology to the plant virus genome (e.g. [4,10,11]). However, it is labour intensive and may be affected by 3′-terminal methylations of small RNAs [10,12]. Presence of viral gene silencing suppressors increases the proportion of unmethylated miRNAs in *Arabidopsis* [13]. The presence of HC-Pro (helper component proteinase) of potyviruses resulted in unmodified 21 and 22-nt viral siRNA populations while the other small RNA species were methylated at 3′-termini [10]. 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 [11], and in *Arabidopsis* [6]. Here, we observed common siRNA hotspots for the 21 and 22-nt species, from TuMV infected *Brassica juncea* (cv. tender green) leaves, providing clear evidence that DCLs processed the TuMV RNA at the same siRNA hotspots.

There is a large amount of information about small RNA strand asymmetry based on the operational characteristics of Argonaute protein [14]. In comparison, 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 GC bias in the siRNAs versus the viral genome, as well as hotspot siRNAs versus non-hotspot siRNAs, 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 [15]. We analyzed approximately 4300 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. In contrast, miRNAs of all 24 vertebrate species displayed a higher AU% than their precursors, indicating a remarkable divergence between plant and animal RNAi machineries.

0014-5793/$32.00
doi:10.1016/j.febslet.2007.06.022

2. Materials and methods

2.1. 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-week after sowing, and were tested by ELISA to confirm systemic TuMV infections 3-week after inoculation [11]. Then young leaves with viral symptoms were sampled and pooled for small RNA cloning and sequencing as previously described by Ho et al. [11]. Leaves from 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).

2.2. Data processing and analyses

Small RNA sequences were excised from the adaptor sequences, compared to the virus genome sequence, and had their polarity determined [11]. Only siRNAs that were 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 (S.E.) were calculated and two-tailed homoscedastic *t*-tests (Microsoft Excel) were performed to compare the GC% among the small RNAs.

3. Results and discussion

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

From the 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, 21 and 22-nt species were enriched (Fig. 1B) and most of them ($n = 595$) were of viral origin with both sense (60%) and antisense (40%) polarity (Fig. 2A). These results were consistent with observations made from HC-Pro expressing tobacco [10], *Arabidopsis* infected by a GFP-tagged-TuMV (TuMV-GFP) [6], and the previous report of a significant level of PTGS of the TuMV dsRNA intermediates [11]. Viral dsRNA intermediates were also implicated as substrates of Dicer-2 in *Drosophila* [16]. Predictions of RNA secondary structure also supported the view of PTGS targeting of TuMV dsRNA because stem–loop structure was not in evidence for all 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 (9798-nt). When plotted using a window of 100-nt, this siRNA population provided an almost complete coverage of the whole virus genome (Fig. 2B).

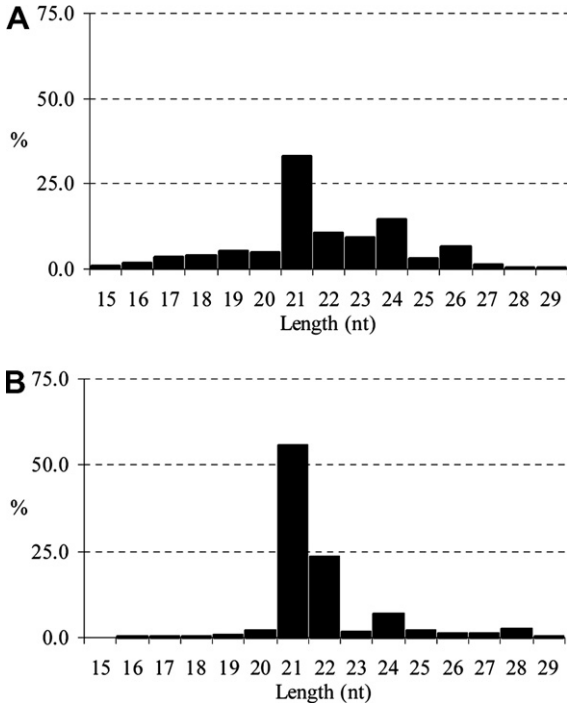

Fig. 1. Size distributions of small RNAs. (A) Sequenced small RNAs ($n = 637$) from Mock (water inoculated) plants. (B) Sequenced small RNAs ($n = 842$) from TuMV infected plants.

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 and 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 60.1–28.2% 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 [7–9]. 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 [7–9]. 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 site despite its overall redundant role to DCL4.

3.2. TuMV siRNA displayed a GC bias

The TuMV siRNA hotspots were strand specific (Fig. 2C), demonstrating Argonaute protein mediated asymmetry [14] in the plant, although the asymmetry of the TuMV siRNAs was apparently not governed by the same rules of animal siRNA production (Ho unpublished data). Analysis on compositional bias of the endogenous small RNAs from the Mock inoculated *B. juncea* sample showed a GC content of

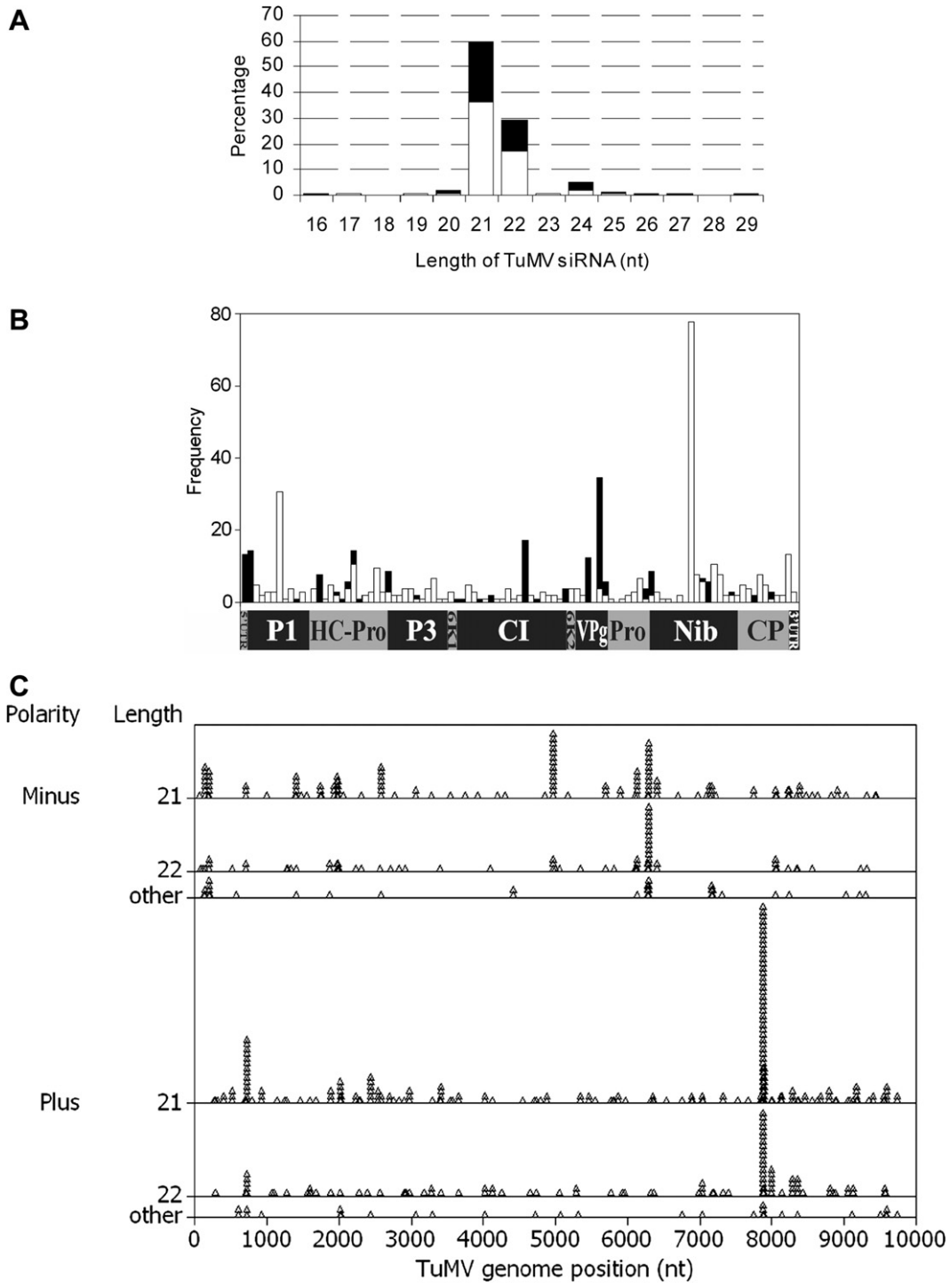

Fig. 2. Polarity, size, and location distributions of TuMV siRNAs (n = 595). (A) Polarity of TuMV-derived siRNAs in each size class. Y-axis shows the percentage of total TuMV siRNAs with a certain length. Open areas shows the proportion of sense siRNA and filled areas represent that of antisense siRNA. (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.

62.3 ± 0.42% (mean ± S.E., n = 637), considerably higher than the GC% = 48.84 based on *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

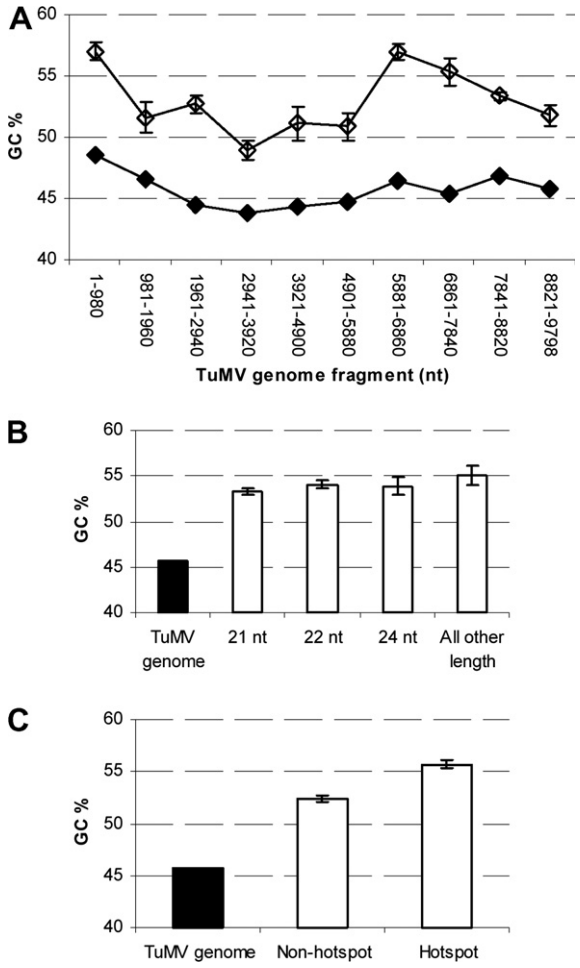

Fig. 3. GC bias of TuMV siRNAs. (A) GC% of TuMV genomic RNA (filled marks) and siRNA (open marks) was calculated for genome fragments of 980-nt long. The error bars represent S.E. (B) Mean GC% of TuMV genomic RNA (filled bar) and siRNA (open bars) of different lengths. The error bars represent SE. (C) GC% of TuMV genomic RNA (filled bar) and siRNA (open bars) from hotspots or non-hotspots. The error bars represent S.E.

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-test) (Fig. 3A). Similar feature was revealed by 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 ($P < 0.0001$, t-test) 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%.

3.3. 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 [15]. 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-test) (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-test) (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 [2]. Additionally, because monocots have evolved high GC usage at their 3rd codon positions (GC_3) [17,18], 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 [19]. As non-conserved miRNA genes may arise from the recent evolutionary past while conserved miRNA genes are considered to be ancient [20–22], we further compared GC% of conserved miRNAs (among *Arabidopsis*, *Populus*, and *Oryza* [20]) 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-test) (Supplementary Fig. 3). In all of these three plant species, GC% of conserved miRNAs were higher than those of non-conserved miRNAs ($P < 0.0001$, t-test) (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

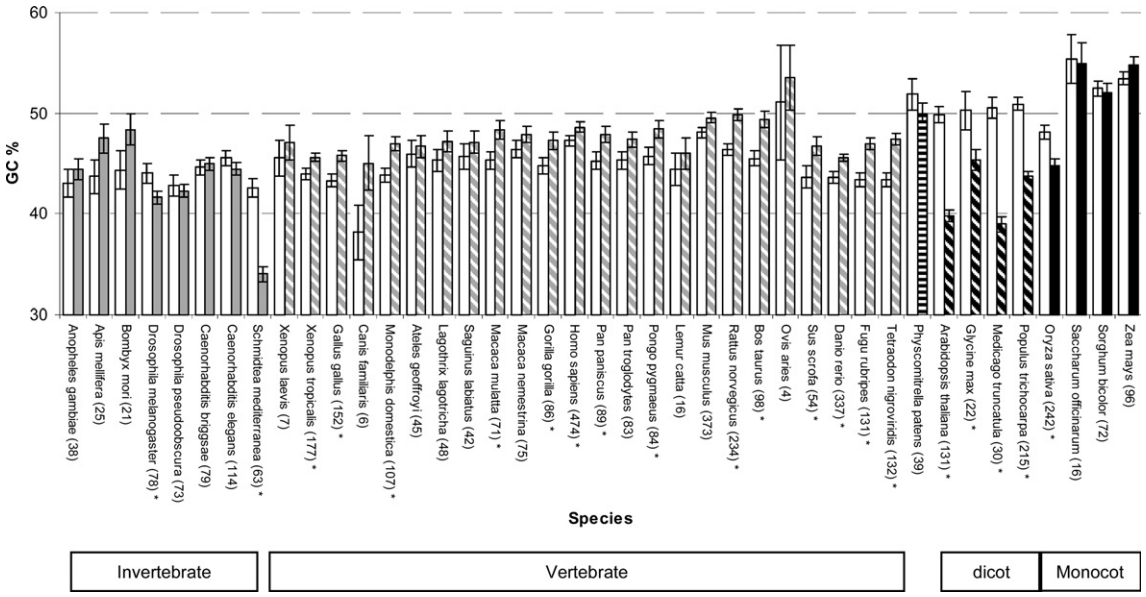

Fig. 4. GC contents of miRNAs. GC% (mean ± S.E.) of miRNAs (open bars) is presented together with their precursors (filled bars) of invertebrate, vertebrate, dicot, and monocot species. The number of miRNA sequences is shown beneath a species name. Asterisk labels significant difference ($P < 0.05$) between miRNA and precursor.

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-test) (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 [23]. 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 and 4), it should be considered, in addition to the strand asymmetry, in applications of RNAi.

Acknowledgements: We are grateful to John Walsh (HRI Warwick, UK) for providing the TuMV isolate and sequence, to Delia McCall for maintaining plants in glasshouse, and to referees who provided thoughtful comments. This work was supported by the Vietnamese Studentship to T.H. (Ministry of Education and Training, Decision No 322/QD-TTg), and NERC (UK) grants to T.D. (NER/A/S/2003/00547) and H.W. (NER/A/S/2003/00548).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2007.06.022.

References

[1] Brodersen, P. and Voinnet, O. (2006) The diversity of RNA silencing pathways in plants. Trends Genet. 22, 268–280.
[2] 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, 2442–2450.
[3] Baulcombe, D. (2004) RNA silencing in plants. Nature 431, 356–363.
[4] 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.
[5] 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. USA 103, 19593–19598.
[6] 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 Biol. 2, e104.
[7] 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.
[8] Blevins, T. et al. (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res. 34, 6233–6246.
[9] Fusaro, A.F. et al. (2006) RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. EMBO Rep. 7, 1168–1175.
[10] 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, 13398–13403.
[11] 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.
[12] 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, 932–935.
[13] 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, 3117–3120.

[14] Hutvagner, G. (2005) Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. FEBS Lett. 579, 5850–5857.
[15] Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297.
[16] 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, 452–454.
[17] 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, 851–856.
[18] Zhou, H., Wang, H., Huang, L.F., Naylor, M. and Clifford, P. (2005) Heterogeneity in codon usages of sobemovirus genes. Arch. Virol. 150, 1591–1605.
[19] 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, 1282–1290.
[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.
[21] 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, 3407–3425.
[22] Fahlgren, N. et al. (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE 2, e219.
[23] Jing, Q. et al. (2005) Involvement of microRNA in AU-rich element-mediated mRNA instability. Cell 120, 623–634.