Indian Journal of Biotechnology
Vol. 17, July 2018, pp 520-526

Comparative study on sequence characteristics of mature and precursor miRNAs of monocot and dicot plants

Chittabrata Mal[1,2] and Sudip Kundu[1*]

[1]Department of Biophysics, Molecular Biology and Bioinformatics, University of Calcutta, 92 APC Road, Kolkata 700009, India

[2]Institute of Biotechnology, Amity University Kolkata, Major Arterial Road (North-South), AA II, Newtown, Rajarhat, West Bengal 700135, India

Received 14 August 2014; revised 11 November 2016; accepted 22 November 2016

MicroRNAs (miRNAs) negatively regulate mRNAs at post-transcriptional level and thus can regulate different biological processes. Here we analyze the sequence characteristics and length distribution of miRNA sequences of six monocot and six dicot plants. We observe a species specific nucleotide preference of miRNAs. Although the first position of 5' end of mature miRNAs is U rich, there exists a wide variation. While the GC% of monocot mature miRNAs in general is nearly equal to the AU%, the GC% of stress induced miRNAs is significantly higher than AU%. Thus, higher content of GC% can be used as a signature of stress induced miRNAs in monocot. The length distribution of mature miRNAs shows the highest peak at 21 nucleotides with some other minor peaks at 20, 22 and 24 nucleotides. The synthesis of some miRNAs has positional preference either to 5' or to 3' arm of their precursors, but they are different in monocot and dicot.

Keywords: Base sequence, mature and precursor miRNA, monocot, dicot, stress target

Introduction

MicroRNAs (miRNAs) are endogenous non-coding RNAs that bind to messenger RNAs (mRNA) and negatively regulate mRNAs at the post-transcriptional level[1]. It can regulate a wide variety of biological processes including transcriptional regulation, development, metabolism and stress responses. The biogenesis process and the main activity of a miRNA are well studied in case of animals as well as plants. Mature miRNAs are single stranded RNAs of about 18-24 nucleotides. They are derived from precursor miRNAs (pre-miRNAs) and these intermediate pre-miRNAs are actually synthesized from longer non-coding primary miRNAs (pri-miRNAs) by the sequential actions of the DCL1 and HYL1 in plants[2]. Mature miRNAs while bind with proteins of the Argonaute (Ago) family, determine the functional output of the associated small RNA (transcriptional silencing, mRNA cleavage or translation repression). One of the most important criteria for annotating a cloned sequence as a miRNA include its characteristic length (approximately 21 nucleotides) and a corresponding compact pre-miRNA loop structure[3,4]. In *Arabidopsis*, the relationship between the biological significance and the sequence length heterogeneity had been recognized for a mature miRNA[5]. There are two different species of mir1 68 - one is of length 21 nucleotide and another is of 22 nucleotides. It is reported that only the 21 nucleotide miR168 species are preferentially stabilized by Argonaute1 (Ago1) in *Arabidopsis*[5]. This result suggests the importance of the sequence length distribution of miRNAs. Furthermore, a single miRNA may have multiple mRNA targets[6] and the targets in turn may involve in different biological processes. While analyzing the human miRNA, Fang *et al* found that the average number of target genes of mature miRNAs is positively correlated with their length[7]. Thus, the lengths of mature miRNAs may have also a relationship with their functions. Determination of the function of miRNAs by identifying their targets is a challenging task. Since the miRNAs recognize their targets through base-pairing, the sequence differences among the related miRNAs are important. Debernardi *et al* showed that miR319 and miR159 having very small differences in their sequences, regulate different genes[8]. Thus, the studies of sequence length and nucleotide distributions of mature miRNAs in the genome are subjects of great interest. Moreover, the base composition is also significant because it is crucial to

*Author for correspondence:
skbmbg@caluniv.ac.in

load miRNAs to the RNA-induced silencing complex (RISC) complex which ultimately binds to the target mRNAs. The biasness in the percentage of nucleotide bases may indicate the presence of structural pattern in miRNAs[7]. It is reported that mature miRNAs do not randomly distribute on the stem-loop hairpins of their precursors[9]. The terminals of mature miRNAs tend to locate near loop structures within 1-6 nucleotides rather than in loops or very far from loop structures (> 6 nucleotides). So, position specific features of the precursor miRNAs, if any, will help to identify them. On the other hand, positional frequency of the nucleotides may have influence in predicting stress induced miRNAs[10]. Plants produce different miRNAs due to stress response and high GC content is proposed as a critical parameter to predict stress induced miRNAs[11].

In summary, sequence length, distribution of bases, frequency of nucleotides at different positions, GC and AU content of mature and precursor miRNAs are important characteristics to determine the structure and function of miRNAs. Such properties of miRNAs are well studied for animals and there is a need of such analysis for plants also. Here, we collected miRNA sequences of six monocot plants (i.e., *Brachipodium distachyon, Hordeum vulgare, Oryza sativa, Sorghum bicolor, Triticum aestivum, Zea mays* which are commonly known as purple false brome, barley, rice, sorghum, wheat and maize, respectively) and six dicot plants (i.e. *Arabidopsis thaliana, Glycine max, Malus domestica, Medicago truncatula, Populus trichocarpa, Solanum tuberosum* which are commonly known as *Arabidopsis*, soyabean, apple, isobgul, poplar, potato, respectively). We analysed base composition, GC and AU content, length distribution and location of mature miRNAs within their precursors for these plants. We studied whether there is any correlation between the length of miRNAs and number of their target genes. The GC% and AU% of the stress-induced mature miRNAs and the position of those miRNAs within their precursors were also analyzed.

Materials and Methods

Data

Mature as well as precursor miRNA sequences of the six monocot plants (i.e. *Brachipodium distachyon, Hordeum vulgare, Oryza sativa, Sorghum bicolor, Triticum aestivum, Zea mays)* and of six dicot plants *(Arabidopsis thaliana, Glycine max, Malus domestica, Medicago truncatula, Populus trichocarpa, Solanum tuberosum*) were retrieved from miRBase[12]. We analyzed only those mature miRNAs whose precursor sequences were available. We retrieved all of the predicted targets (expectation cut-off value = 2.0) from a web tool psRNATarget[13]. Stress-induced miRNAs were collected from PASmiR database[14]. Finally, we analyzed only those stress-induced miRNAs whose sequences are available in the miRBase.

Data Analysis by In-house PERL Script

Sequence compositions, the location of mature miRNAs within the precursors, GC% and AU% of mature and precursor miRNAs; relationship between length of the pre-miRNAs and their corresponding mature miRNAs; mature miRNA sequence length and the number of predicted target genes were analyzed by in-house PERL scripts. Sequence logos were generated using Weblogo[15]. Mann Whitney U test was performed by Past3 software[16] to identify whether the two populations are significantly different from each other or not.

In the result section, The p stands for probability and measures how likely it is that any observed difference between groups is due to chance[17].

Results & Discussion

Distribution of Bases in Mature miRNAs

Mature miRNAs were collected for six monocot and six dicot plants. All these plants are economically important and they have the largest number of known miRNAs. Here, we calculated the percentage of four bases present in mature miRNAs of different plants. Among the monocot plants, both the percentage of pyrimidine residues (C and U) are predominant in maize ($p < 2.28E-05$, $p < 0.008$); while the percentage of purine residues A and G are predominant in sorghum ($p < 0.045$) and barley ($p < 0.037$), respectively. Among the dicot plants, the percentage of pyrimidine residues C and U are predominant in apple ($p < 1.46E-19$) and isobgul ($p < 1.96E-11$), respectively; while the percentage of purine residues A and G are predominant in potato ($p < 0.024$) and apple ($p < 0.048$), respectively. From the Figure 1 it could be stated that (i) maize has higher C% than the averages of other five monocots ($p < 2.28E-05$) and six dicots ($p < 7.22E-06$); and (ii) maize has also higher U% than the averages of other five monocots ($p < 0.008$). Again we found that the monocot mature miRNAs are GU-rich, where G residues are more frequent than other bases (data not shown). In case of

dicot, mature miRNAs are AU-rich where U residues are more frequent than other bases. Thus, the results indicate that there is a significance of specie-specific base preference in miRNAs.

Position Specific Base Composition Characteristics of miRNAs

Zhang *et al* showed earlier that the U residue is dominant in the first position of 5' end of *Arabidopsis* and rice mature miRNAs[18]. Here we extensively studied position specific base variation for six mononcot and six dicot plants using the updated and experimentally identified matured miRNAs from miRBase[12]. We also observed that the U is predominant than the others at the first position of 5' end of mature miRNAs in all plants (Fig. 2). However, a wide range of variation of relative occurrence of U at that position in monocot and dicot (45% to 87% and 45% to 69%, respectively) is observed and the values are listed in the Table 1. Further, the targets of plant mature miRNAs are cleaved by Ago1 protein complex at a specific position, opposite to the 10th and 11th nucleotides of the miRNA[19]. So, we analyzed the 10th and 11th position of mature miRNAs to find whether there is any nucleotide preference at those positions. Here, we observed that the nucleotide preference at 10th

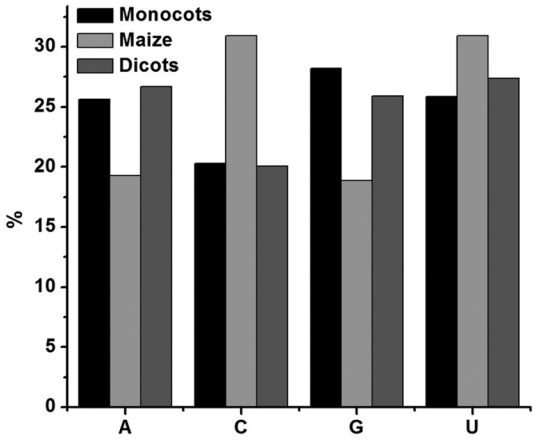

Fig. 1 — Nucleotide distribution of mature miRNAs of the five monocot plants, maize and six dicot plants.

Table 1 — Percentage of U at 1st position of 5' end of mature miRNA

	Plant species	Percentage of U
Monocot	*B. distachyon*	72.39
	H. vulgare	49.23
	O. sativa	45.51
	S. bicolor	57.64
	T. aestivum	57.14
	Z. mays	86.95
Dicot	*A. thaliana*	65.44
	G. max	55.05
	M. domestica	68.93
	M. truncatula	45.68
	P. trichocarpa	66.76
	S. tuberosum	38.39

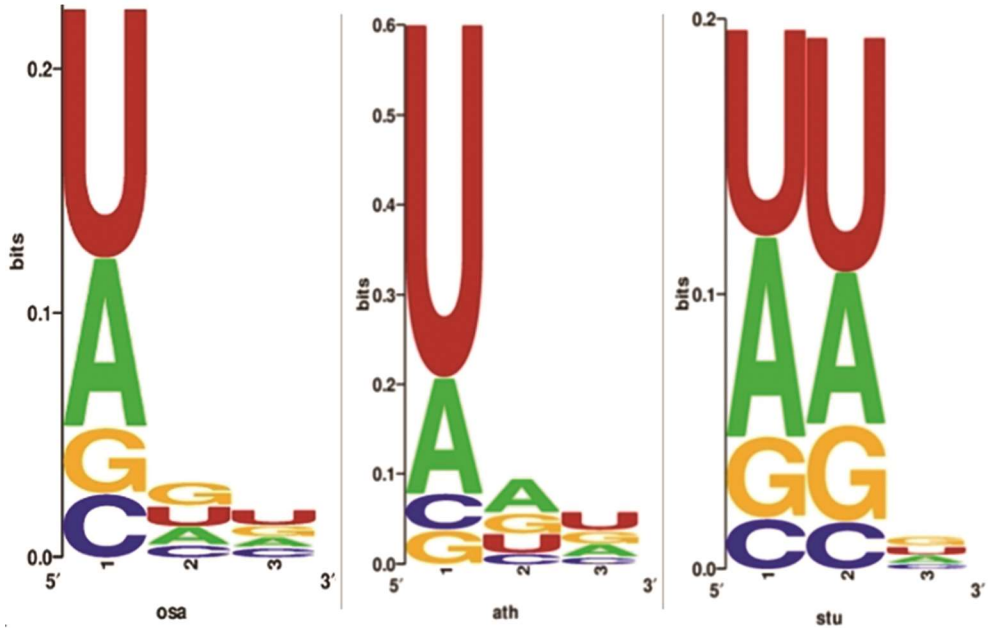

Fig. 2 — Frequency of the 1st, 10th and 11th bases at 5' end of mature miRNAs of three plants. In the X axis 1, 2, 3 indicate 1st, 10th and 11th bases, respectively. Abbreviations- osa: O*ryza sativa* (rice, monocot), ath: *Arabidopsis thaliana* (dicot), stu: *Solanum tuberosum* (potato, dicot).

position of potato is similar to the 1st position; however there is a slight variation in their relative distribution (Fig. 2). But, no significant difference in base composition at 10th and 11th position of miRNAs was observed in other plants. Thus it can be concluded that nucleotide distribution of 10th position of mature miRNAs of potato plant may bear some potential significance relating the binding preference of miRNAs to their targets.

GC% and AU% of Mature and Pre-miRNAs

Next we compared GC and AU% of mature and pre-miRNAs of monocot (Fig. 3a) and dicot (Fig. 3b) plants. Fang *et al* showed that the GC% is very close to the AU% in human mature miRNAs[7]. We observed the same trend in mature miRNAs of monocot plants, i.e., GC% is nearly equal to the AU% (no statistical significant difference). However, AU% is significantly higher than the GC% of dicot mature miRNAs ($p < 0.008$) (Fig. 3b). The AU% is significantly higher than the GC% of both the monocot and dicot pre-miRNAs ($p < 0.008$ and $p < 0.005$, respectively). However, in dicot plants, the GC% and AU% of mature miRNAs are significantly ($p < 0.01$) higher and lower than those of the precursor miRNAs (Fig. 3b). Although it is evident from Figure 3a that the GC% and AU% of mature miRNAs in monocot plants are higher and lower, respectively than those of the precursor miRNAs, the differences are not statistically significant. We observed an exception that maize (a monocot) mature miRNAs have lower frequencies of GC% than precursor miRNAs.

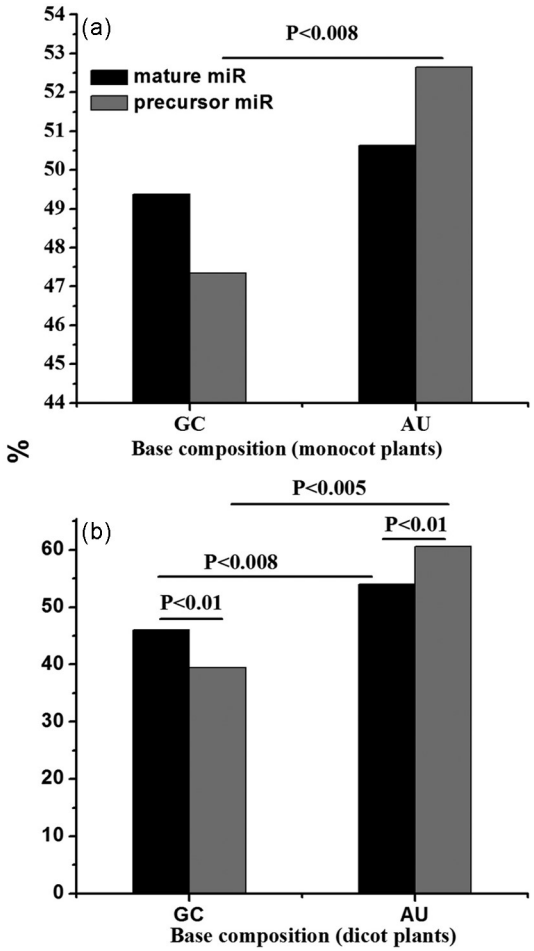

Fig. 3 — GC% and AU% of mature and premature sequence: (a) GC% and AU% of mature and premature sequence of monocot plants and (b) GC% and AU% of mature and premature sequence of dicot plants.

Mishra *et al* showed that stress regulated miRNAs in *A. thaliana* have higher GC%[11]. As stress-induced miRNA data was unavailable for sorghum and apple plants, here we analyzed GC% and AU% for the rest ten plant species. We also observed that the stress-induced mature miRNAs of monocot plants have significantly higher GC% than AU% ($p < 0.01$) (Fig. 4a). But, in case of dicot plants, the AU% is nearly equal to the GC% (Fig. 4b). The GC% of precursor miRNAs of stress induced monocot plants is significantly higher than the AU% ($p < 0.02$), while AU% is higher than GC% in the stress-induced precursor miRNAs of dicot plants ($p < 0.01$).

Distribution of the Lengths of the Pre-miRNAs and their Corresponding Mature miRNAs

It is already known that one of the most important criteria used to annotate a cloned sequence as a miRNA is its characteristic length (nearly 22 nucleotide)[4]. However, Fang *et al* showed that the length of human mature miRNAs varies from 16 to 27 with a median of 22 and that of pre-cursors ranges from 41 to 188[7]. Here, in case of plants, the lengths of mature miRNAs vary from 18 to 26, and all the plants have higher frequency of mature miRNAs of 21 nucleotides. However, most of the plants also have considerable frequencies of mature miRNAs with length of 20, 22 and 24 nucleotides (Fig 5a & 5b). The sequence lengths of pre-miRNAs vary from 59 to 541 in monocot and 43 to 918 in dicot. The distribution of the sequence lengths of pre-miRNAs of the plants showed that most of the plants have similar kind of distribution pattern with a right skew. For example, the distribution pattern for rice (monocot) and *Arabidopsis* (dicot) are shown in the Figure 6a & 6b, respectively.

Then we investigated the relationship between the sequence lengths of mature miRNAs and the

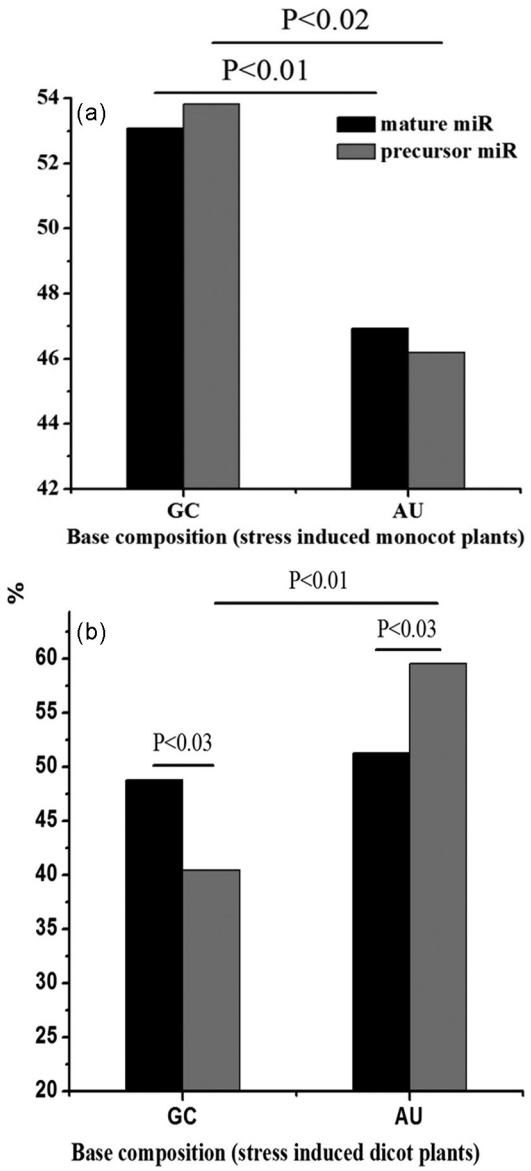

Fig. 4 — GC% and AU% of stress-induced mature and precursor miRNAs: (a) GC% and AU% of stress-induced mature and precursor miRNAs of monocot plants and (b) GC% and AU% of stress-induced mature and precursor miRNAs of dicot plants.

corresponding pre-miRNAs. There are multiple pre-miRNA lengths corresponding to each mature miRNA length. In case of human, Fang *et al* found that there is a positive, significant relationship between the mature and precursor miRNA sequence lengths[7]. However, we did not observed any significant relationship between the length of the mature and precursor miRNAs of all the plants.

Location of Mature miRNAs within the Precursors

Here, we tried to identify whether the mature miRNAs have any positional preference while

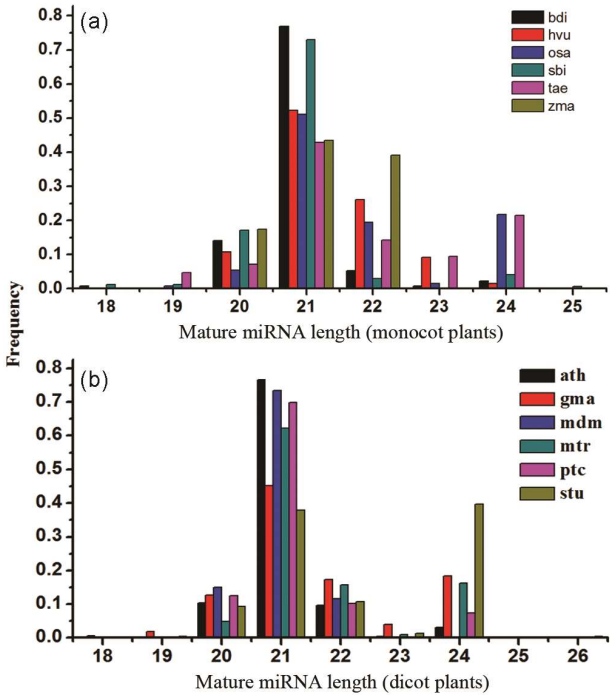

Fig. 5 — Mature miRNA length distribution: (a) Mature miRNA length distribution of monocot plants and (b) Mature miRNA length distribution of dicot plants.

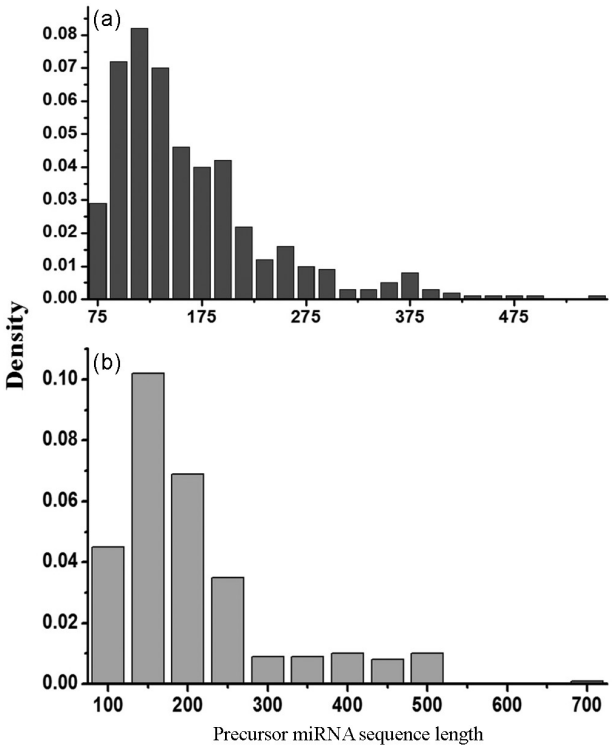

Fig. 6 — Distribution of the sequence length of pre-miNRAs: (a) Distribution of the sequence length of rice (monocot) pre-miRNAs and (b) Distribution of the sequence length of *Arabidopsis* (dicot) pre-miRNAs.

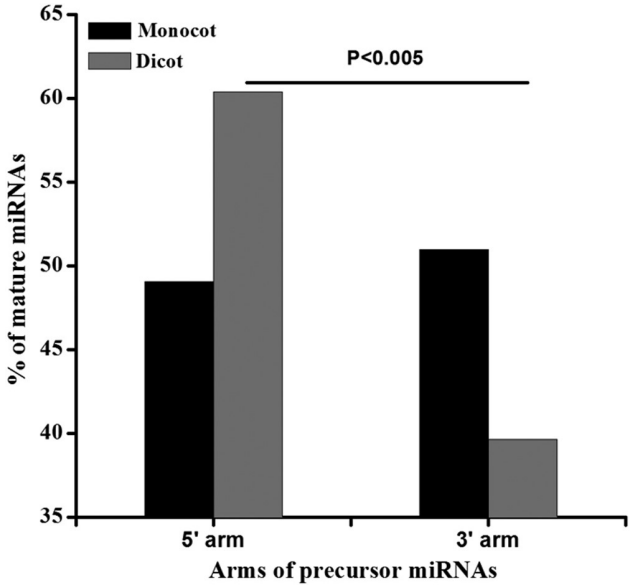

Fig. 7 — Distribution of monocot and dicot mature miRNAs in the 5' and 3' arms of the precursors.

synthesized from the precursor miRNAs. Some miRNAs do not show any preference towards any particular arm (5' or 3') of the precursors. We further analyzed only those miRNAs which have a positional preference towards any arm. In case of rice, sorghum, and maize, the higher numbers of mature miRNAs are synthesized from 3' arm of the precursors and in case of purple false brome, barley and wheat, higher numbers of mature miRNAs are synthesized from 5' arm of the precursors (data not shown). In case of dicot plants, higher numbers of mature miRNAs are synthesized from 5' arm of the precursors ($p < 0.005$) (Fig. 7).

Mature miRNA Sequence Length and the Number of Predicted Target Genes

MicroRNA or miRNAs bind to the target mRNAs and thus destabilize the mRNAs or suppress its translation. The non-coding but functional miRNAs may have multiple target genes. We retrieved all of the predicted targets (expectation cut-off value = 2.0) from a web tool psRNATarget[13]. Fang *et al* showed that the average number of the target genes is positively correlated to the sequence length of the miRNA in human[7]. Here, we did not observed any significant correlation for any of the six monocot and six dicot plants and thus probably indicating that the relationship is different in plants from humans.

In conclusion, we studied the sequence and length characteristics of miRNAs of 12 economically important plants (six monocots and six dicots). The results indicate the species specific nucleotide preferences for some plants. Pyrimidine residues are pre-dominant in maize; purine residues are predominant in sorghum and barley, respectively for mature miRNAs. Among the dicot plants, the percentage of pyrimidine residues C and U are predominant in apple and isobgul, respectively; while the percentage of purine residues A and G are predominant in potato and apple, respectively. Although U is predominant than the others at the first position of 5' end of mature miRNA in all plants, there are variation in the amount of U at that position within the plants. In general, the AU% is significantly higher than the GC% of both the monocot and dicot pre-miRNAs. However, in case of stress-induced monocot mature miRNAs, the GC% is higher than the AU%, while AU% is higher than GC% in those of dicot plants. Thus, the higher GC% of miRNAs can be used as an indicator for finding monocot stress-induced miRNAs. But, the same could not be used for dicot plants. While the majority of mature miRNAs in rice, sorghum and maize are synthesized from 3' arm of the precursors, majority of those in purple false brome, barley and wheat are synthesized from 5' arm. On the other hand, higher number of mature miRNAs is synthesized from 5' arm of the precursors in the dicot plants. All of these characteristics features could be useful for miRNA annotation as well as target prediction.

Acknowledgements

CM gratefully acknowledges University Grants Commission (Research Fellowship for Meritorious Students in Science) and DST-PURSE, Government of India for financial support and Distributed Information Center, University of Calcutta for computational facilities. SK acknowledges Centre of Excellence in Systems Biology and Biomedical Engineering for partial support.

References

1 Zhang B, Wang Q & Pan X, MicroRNAs and their regulatory roles in animals and plants, *J Cell Physiol*, 210 (2007) 279-89.

2 Voinnet O, Origin, biogenesis and activity of plant microRNAs, *Cell*, 136 (2009) 669–687.

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

4 Ambros V, Bartel B, Bartel D P, Burge C B, Carrington J C *et al,* A uniform system for microRNA annotation, *RNA*, 9 (2003) 277–279.

5 Vaucheret H, Ago1 homeostasis involves differential production of 21-nt and 22-nt miR168 species by MIR168a and MIR168b, *PLoS ONE,* 4 (2009) e6442.

6 Jones-Rhoades M W & Bartel D P, Computational identification of plant microRNAs and their targets, including a stress-induced miRNA, *Mol Cell,* 14 (2004) 787–799.

7 Fang Z, Du R, Edwards A, Flemington E K & Zhang K, The sequence structures of human microRNA molecules and their implications, *PLoS ONE,* 8 (2013) e54215.

8 Debernardi J M, Rodriguez R E, Mecchia M A & Palatnik J F, Functional specialization of the plant miR396 regulatory network through distinct microRNA target interactions, *PLoS Genetics,* 8 (2012) e1002419.

9 Huiyu X, Fei L, Tao H & Yanda L, Distribution of mature microRNA on its precursor: A new character for microRNA prediction, *Inter J Infor Technol,* 11 (2005).

10 Gupta R, Soni N, Patnaik P, Sood I, Singh R *et al*, High AU content: a signature of upregulated miRNA in cardiac diseases, *Bioinformation,* 5 (2010) 132–135.

11 Mishra A K, Agarwal S, Jain C K & Rani V, High GC content: Critical parameter for predicting stress regulated miRNAs in *Arabidopsis thaliana*, *Bioinformation,* 4 (2009) 151–154.

12 Kozomara A & Griffiths-Jones S, miRBase: integrating microRNA annotation and deep-sequencing data, *Nucleic Acids Research,* 39 (2011) D152–D157.

13 Dai X & Zhao P X, psRNATarget: A plant small RNA target analysis server, *Nucleic Acids Research,* 39 (2011) 155–9.

14 Zhang S, Yue Y, Sheng L, Wu Y, Fan G *et al*, PASmiR: a literature curated database for miRNA molecular regulation in plant response to abiotic stress, *BMC Plant Biol,* 13 (2013) 33.

15 Crooks G E, Hon G, Chandonia J M & Brenner S E, Weblogo: A sequence logo generator, *Genome Res*, 14 (2004) 1188–1190.

16 Hammer, Harper D A T & Ryan P D, PAST- Paleontological Statistics Software Package for education and data analysis, *Paleontologia Electronica* 4 (2001) 1–9.

17 Goodman S N, p values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate, *Am J Epidemiol,* 137(5) (1993) 485-96.

18 Zhang B, Pan X, Cannon C H, Cobb G P & Anderson T A, Conservation and divergence of plant microRNA genes, *Plant,* 46 (2006) 243–259.

19 Mallory A C, Elmayan T & Vaucheret H, MicroRNA maturation and action–the expanding roles of ARGONAUTEs, *Curr Opin Plant Biol*, 11 (2008) 560–6.