Journal of Computational Intelligence in Bioinformatics
ISSN: 0973-385X Volume 2 Number 2 (2009) pp. 109–119

http://www.ripublication.com/jcib.htm

Computational Identification of MicroRNA Homologs from *Acyrthosiphon pisum* (Pea Aphid)

Ganesh Sathyamurthy[1*] and N. Ramachandra Swamy[1]

[1]Department of Biochemistry, Bangalore University, Bangalore-560 001, India
Corresponding author e-mail: [1]ganeshsmurthy@msn.com; [1]sganesh@bub.ernet.in

Abstract

MicroRNAs (miRNAs) are a large class of small non coding miRNAs that regulate gene expression. In this paper we use the comparative miRNA method for computational identification miRNAs from *Acyrthosiphon pisum*, Pea aphid, belonging to order Hemiptera and superfamily Aphidoidea. We validate the newly identified miRNAs using MiPred for the classification of real and pseudo miRNAs. We thus report 10 newly identified miRNAs, of which 9 are real precursors and mir-1923 being pseudo precursor. 1 of 10 miRNAs has 96.43% identity, 4 between 90 – 94.9% identity, 2 between 85-89.9% identity and 3 of less than 85% identity percent. The statistical sequence characteristics are defined for all 10 miRNAs. We further observe that U being a predominant at end base in mature sequence. We conclude thus, our study to be the beginning of understanding the *A. pisum* genome and may play an important role in effective pest management in near future.

Running title: Computational Identification of miRNAs in Acyrthosiphon pisum

Key words: Invertebrates, Insect, Base frequency, real miRNA, pseudo miRNA.

Introduction

MicroRNAs (miRNAs), approximately 22 nucleotide (nt) long, are a large class of small non-coding RNAs (ncRNAs), which play an important role in control of gene expression at post transcriptional level (Rodriguez, Griffiths-Jones et al. 2004) (Fanjul-Fernandez, Folgueras et al. 2009). Many studies reveal that multicellular eukaryotes use miRNAs to regulate basic cellular functions including proliferation, differentiation and death (Schaar, Medina et al. 2009) (Yan, Zhou et al. 2009) (le

Sage, Nagel et al. 2007) (Xia, He et al. 2009) (Chen, Olsen et al. 2009). Most of miRNAs are phylogenetically conserved; however, there are also many non-conserved, species-specific miRNAs that might have a role in the emergence of phenotypic variation in closely related species (Yao, Zhao et al. 2009) (Barakat, Wall et al. 2007) (Li, Tang et al. 2007). Some miRNAs are present in the genome as clusters where two or more miRNAs occupy neighboring position within a few kilobases of each other and are transcribed as polycistronic structures. These polycistronic miRNAs may share cooperative function (Thatcher, Bond et al. 2008) (Zhou, Wang et al. 2009) (Xu, Witmer et al. 2007). In animals, these miRNAs target genes through miRNA-complementary sites located at the 3` untranslated regions (UTRs). Most animal miRNAs bind with mismatches and bulges, although a key feature of recognition involves Watson-Crick base pairing of miRNA nucleotides 2-8, representing the seed region (Cora, Di Cunto et al. 2007) (Ruan, Chen et al. 2008) (Yang, Wang et al. 2008). In contrast, most plant miRNAs bind with near perfect complementarily to sites within the coding sequence of their targets (Yang, Wang et al. 2008). This difference in degree of complementarity is considered as a key determinant of regulatory mechanism (Carthew and Sontheimer 2009).

By March 2009, 9539 miRNA sequence entries have been registered in miRBase (Griffiths-Jones, Grocock et al. 2006) (Griffiths-Jones 2004) (Ambros, Bartel et al. 2003). Sequence analysis have shown that some mature miRNAs are phylogenetically conserved, particularly in the atleast first 6-8 residues at 5` end in species of the same kingdom (Griffiths-Jones, Grocock et al. 2006). Quite a few mature miRNA sequences are conserved between animals and plants. For example, mir-854 is identified in C. elegans, mouse, human and plants (Arteaga-Vazquez, Caballero-Perez et al. 2006).

Currently there are three methods for identifying miRNAs: the classic cloning method, deep sequencing method and computational approach. The classic and deep sequencing methods are not as efficient as compared to computational approach to detect miRNAs, yet they are more powerful than computational approach to validate the miRNAs. The computational approach could be further categorized into three types: *ab initio* prediction based on the sequence and structure features, comparative genomic approach based on conservation and integrated approach. Presently, majority of the known miRNAs are detected by computational approaches in diverse organisms from plants to higher animals (Joung and Fei 2009) (Watanabe, Tomita et al. 2007) (Adai, Johnson et al. 2005) (Adai, Johnson et al. 2005) (Xie, Huang et al. 2007) (Zhou, Wang et al. 2009).

Till today, there are no reports of miRNAs from *Acyrthosiphon pisum* (*A. pisum*), also referred as Pea aphid. Aphid belongs to order Hemiptera and superfamily Aphidoidea. Currently there are no reports of miRNAs from superfamily Aphidoidea. In the present paper, we take the initiative of reporting novel miRNAs from *A. pisum* by using computational approach of comparative genomic methods. We have examined the statistical characteristics of miRNA genes of all known invertebrate miRNA genes and the novel miRNAs from *A. pisum*.

Materials and Methods

We obtained all the known animal pre- and mature miRNA sequences from Welcome Trust Sanger Institute's, miRBase, release 13, (Griffiths-Jones 2004). The sequences were obtained and stored in multiFASTA format for further analysis. The *Acyrthosiphon pisum* genome Build 1.1 (Sabater-Munoz, Legeai et al. 2006) was used from the NCBI's Genome.

The potential miRNA gene search was performed using BLAST tool (Altschul, Gish et al. 1990) against *Acyrthosiphon pisum*, using published miRNA genes of all invertebrates. The criteria for considering and reporting *A. pisum* pre-miRNA genes were by removing the duplicate miR ID (of different query organism) with low identity percent, bit score value and higher gap values. The remaining sequences were screened for a minimum of 80% identity values with their respective query sequences. We further validated the putative hits of *A. pisum* pre-miRNAs for real and pseudo miRNA genes using MiPred (Jiang, Wu et al. 2007).

A 2 dimensional de novo analysis was performed for both query and hit sequences using RNA shapes (Steffen, Voss et al. 2006). The miR gene sequences of both query and hit sequences were fed into the tool as multiFASTA format and results were obtained as Post script file and used further for visualize the secondary structure (Steffen, Voss et al. 2006).

To analyze the conservation of mature sequence in the miRNA genes, a local alignment was performed, by comparing mature miRNA sequences of query and putative miR gene sequences from *Acyrthosiphon pisum*. We also looked into the sequence statistical characteristics of the miRNA genes, by comparing with the all other known invertebrates miRNA sequences (Table 1). The whole genome sequence (WGS) of putative pre-miRNA sequences of *Acyrthosiphon pisum* were used to identify the genomic locations. The genomic regions were identified and reported (Supplementary).

Organism	Character	length	A %	U %	G %	C %	A+U%	G+C%
Aphid	minimal	62.00	23.94	30.14	17.74	13.24	55.68	33.80
	maximal	88.00	30.14	42.25	24.10	21.92	66.20	44.32
	median	75.00	26.63	31.52	20.68	19.56	60.02	39.98
	mean	75.50	26.77	33.70	20.68	18.85	60.46	39.54
	SD	9.08	1.68	4.20	2.27	2.72	3.63	3.63
All other invertebrates	minimal	49.00	11.88	15.69	9.68	7.69	31.40	19.32
	maximal	215.00	44.29	46.15	40.00	37.21	80.68	68.60
	median	57.33	25.62	31.71	22.50	20.17	57.33	42.67
	mean	90.00	25.00	31.52	22.33	20.00	57.73	42.27
	SD	13.82	4.58	4.91	3.80	4.29	6.85	6.85

Table 1: Comparative statistical sequence characteristic of aphid and other invertebrates.

Results and Discussion

Pre-miRNAs

We report 10 pre-miRNA genes from *A. pisum.* 1 pre-miRNA with 96.43% identity, 4 between 90 – 94.9% identity, 2 between 85-89.9% identity and 3 of less than 85% identity percent. We did not identify any putative pre-miRNAs with 100% identity indicating the divergence in evolution from other closer relatives (insects). This clearly indicates those miRNAs identified here are all being more species specific. Pre-miRNAs tend to form a stem-loop hairpin in their secondary structure. It was previously shown stem-loop structure is not a unique characteristics of miRNAs (Ambros, Bartel et al. 2003). However, in our observation we can classify this into two categories where in some miRNAs there is no much great effect on secondary structures of pre-miRNAs (Figure 1), while in the other there are some differences in the secondary structure (data not shown). The prediction values of MiPred values are shown in Supplementary. Except of miR-1923, all other newly predicted miRNAs are Real precursors, miR-1923 has the prediction percent of 73.6%.

Mature sequences

Of these 10, miRNA genes, 8 of them had 100% identity at mature sequence level. And 2 miRNAs have 92.59% and 95.45% identity percent (Figure 2). This gives clear evidence that the mature sequences have been left untouched through their evolution although the pre-miRNA sequences have not been conserved (Figure 3). This again conserves the functional domains of the miRNAs and their effect on their target genes.

Phylogeny

We were further interested to see the relationship between the query sequences and the putative miRNA hits in the *A. pisum*. We find that all the query miRNAs for the putative hits were from insects alone and not from any other invertebrates, thus indicating the miRNAs are more conserved among closely related organisms. A matrix was constructed with the distribution of putative miRNA hits across the invertebrate species (Figure 4A). Interestingly we find that 3 miRNAs come from *Tribolium*, 2 from *Anopheles*, 2 from *Apis*, 2 from *Bombyx*, 1 from *Drosphila*. Thus accounting for 10 newly identified miRNAs. *A. pisum* belongs to Hemiptera and *Tribolium* to coleoptera one of the closer order in among insects. Further we also constructed a matrix in a similar method for mature sequence distribution with identity percent and query insects used (Figure 4B). Here we find that majority of the sequence fall within the 100% identity percent category indicating the conservation of mature sequences and hence their functionality on their target genes.

Statistical sequence characteristic

Sequence characteristics of pre-miRNAs and mature miRNAs in plants and animals have been reported (Zhou, Wang et al. 2009) (Zhang, Pan et al. 2008) (Dezulian, Remmert et al. 2006) (Berezikov, Guryev et al. 2005). Therefore we looked upon the sequence characteristics of pre-miRNAs and mature miRNAs of *A. pisum* and all other invertebrates. In our search length of miRNA genes of *A. pisum* varied from 62

to 88-nt with a mean of 75.5-nt length. The frequency of bases, A, U, G and C was found to be quite consistent the values of all other known invertebrates, although the 100% identity at pre-miRNA level was not found (Table 1). We find that in the genome of *A. pisum* the length of the mature sequences range between 21 to 27-nt, 22-nt being more common. Recent studies of miRNAs in animal show composition of A+U are greater than G+C in the mature sequences (He, Nie et al. 2008) (Yue, Sheng et al. 2008). Our studies further support with the similar observation wherein, A+G and G+C are 57.94% and 42.06% respectively (Figure 5). It was previously proposed that U is the predominant nucleotide at 5` end of mature sequence in plants thereby playing an important role in their biogenesis through the recognition of targeted miRNA precursor by RNA induced silencing complex (Chan, Ramaswamy et al.) (Zhang, Pan et al. 2006). Among the 10 mature sequences analyzed, we find that U (50%) occurs more than other bases at the 5` end (Figure 6), which was again consistent.

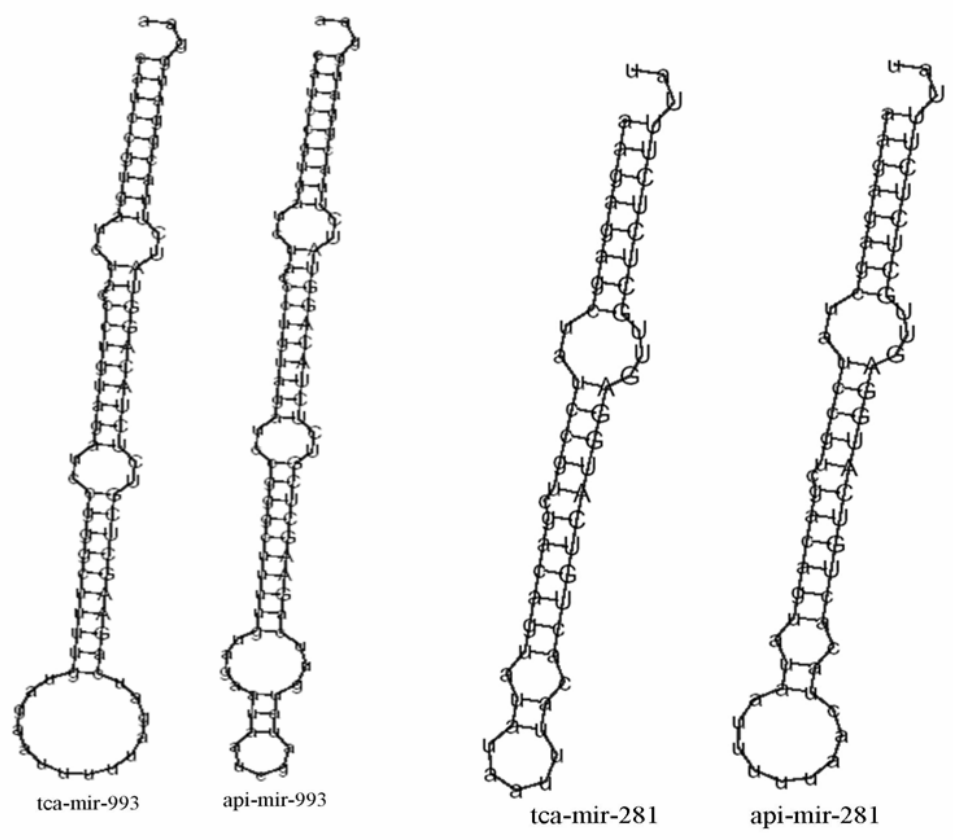

Figure 1: Comparison of secondary structure between *Acrythosiphon pisum* and respective query sequences. A. tca-mir-993 and api-mir-993, B. tca-mir-281 and api-mir-281

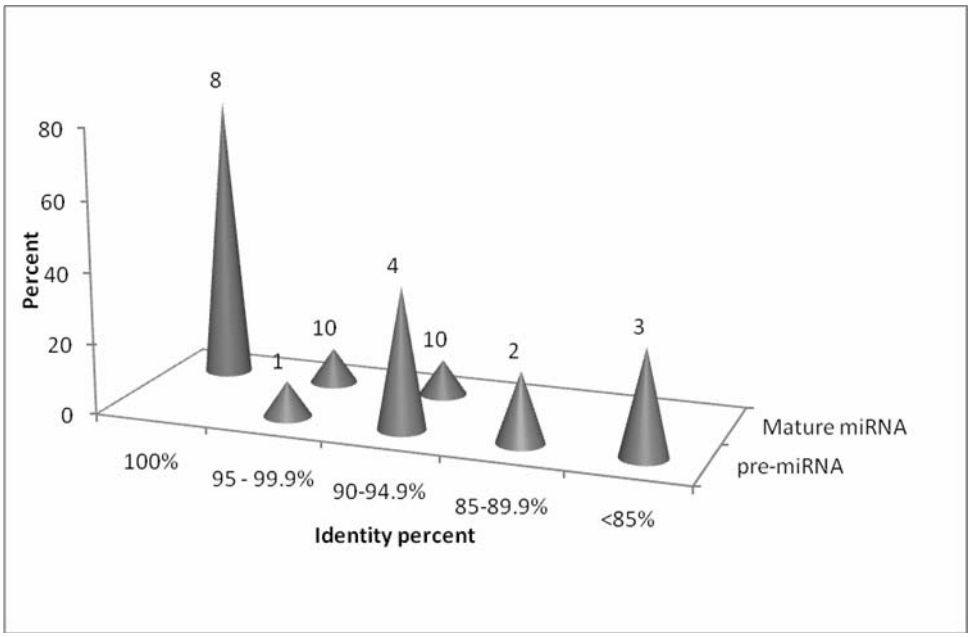

Figure 2: Graph showing counts of pre-miRNA and mature miRNAs.

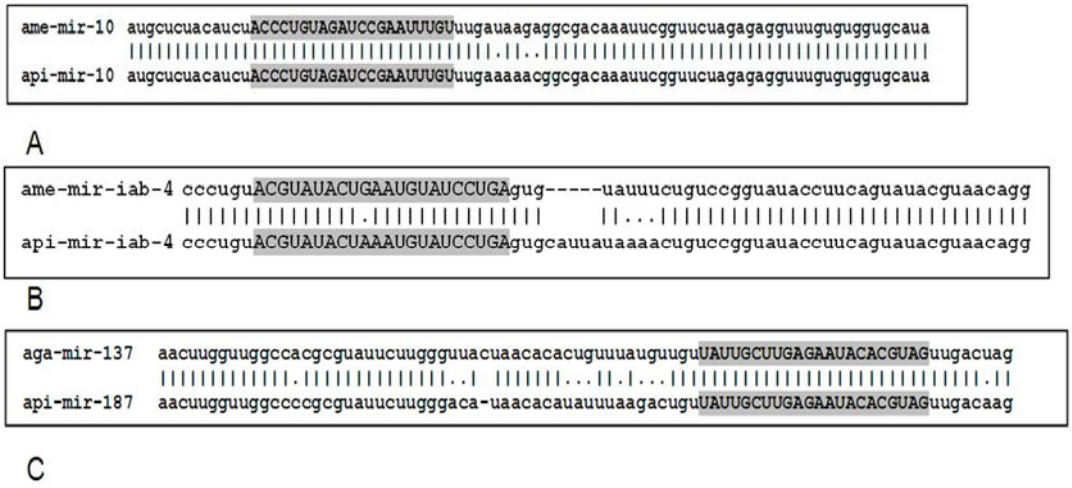

Figure 3: Pairwise alignment comparison of pre- and mature miRNAs of query and aphid sequences. A. ame-mir-10 and api-mir-10 complete conservation of pre-miRNA and mature sequence. B. ame-mir-iab-4 and api-mir-iab-4 having 90.41% and 95.45% identity between pre-miRNA and mature sequences respectively. C. aga-mir-137 and api-mir-187 having 88.24% and 100% identity between pre-miRNA and mature sequences respectively.

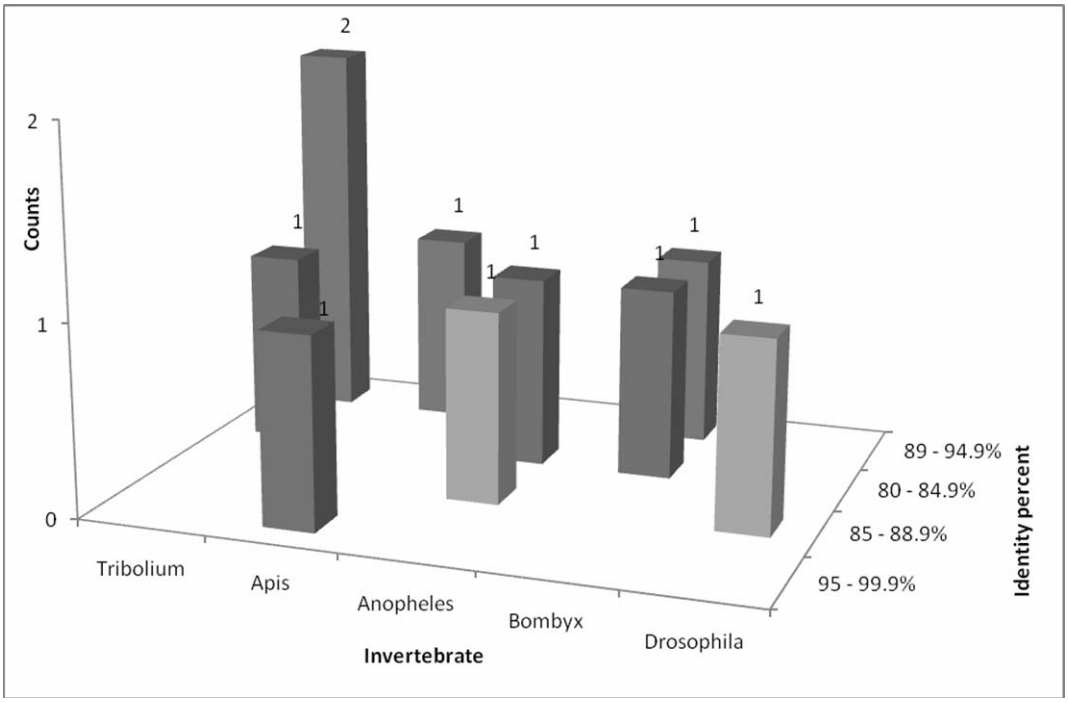

Figure 4A: Histogram showing number of pre-miRNAs distributed across identity percent and query invertebrates.

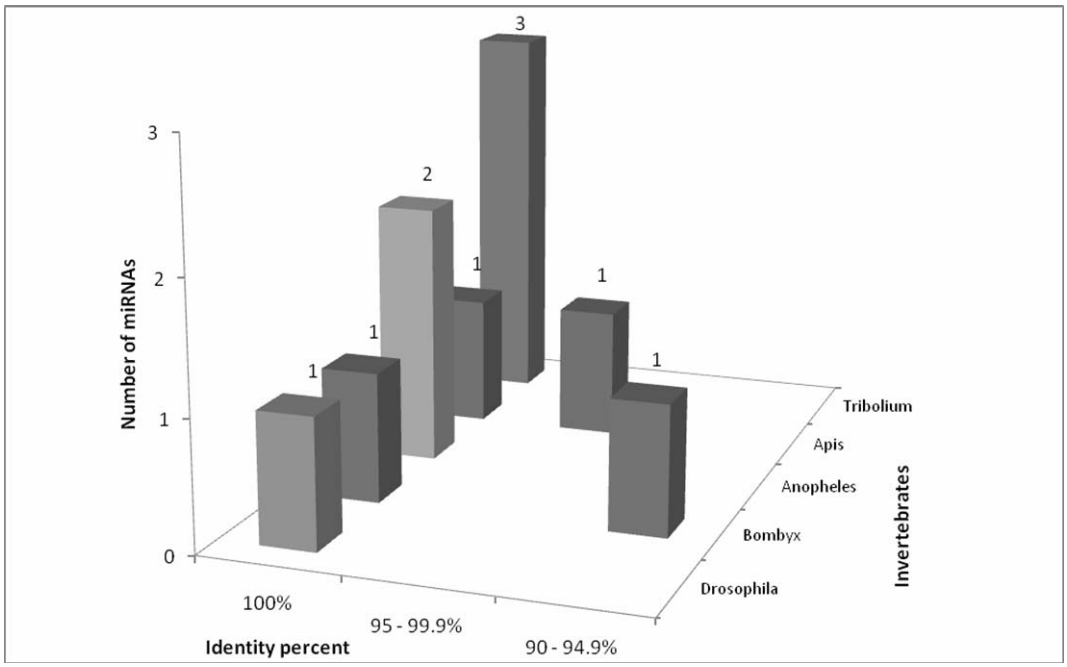

Figure 4B: Histogram showing number of mature miRNAs distributed across identity percent and query invertebrates.

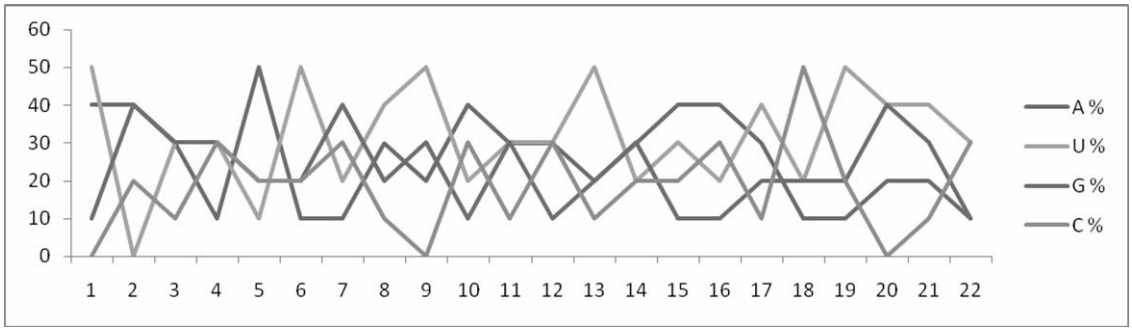

Figure 5: Graph showing base frequency across the length of the mature miRNA sequences. A, U, G, C frequency (%).

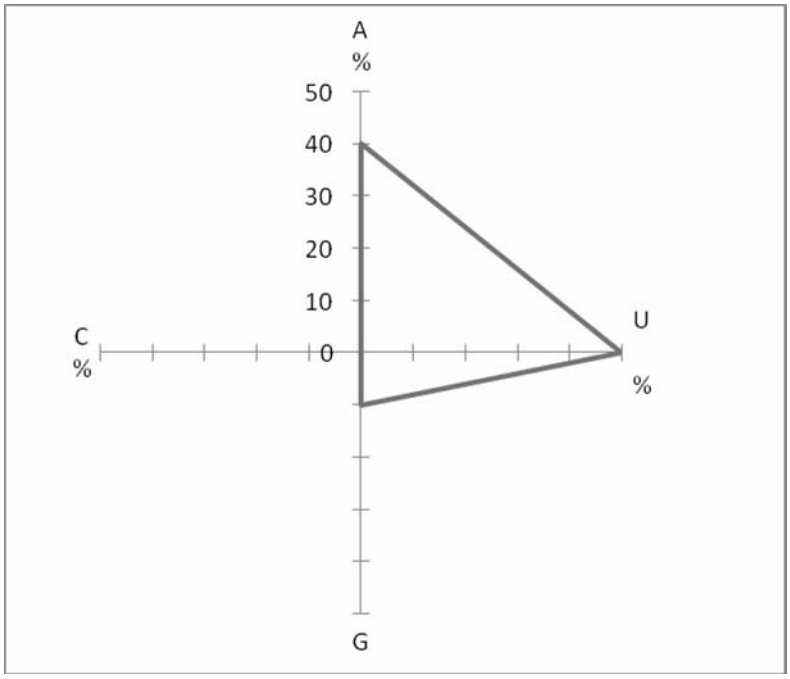

Figure 6: End base frequency in miRNA of cats. Graph showing the occurrence of different nucleotides at 5' terminus.

Conclusion

In our study, we report 10 miRNAs by using miPred and comparative homology using all known invertebrate miRNAs. The statistical sequence characteristics are found to be consistent with all other known invertebrate miRNA sequences. As we have used the comparative method for detecting miRNAs, there could be many more miRNAs yet to be identified in the genome, which could have arised after speciation and remains unidentified in our analysis. miRNA cluster are defined as miRNAs present in the same direction and are transcribed in polycistronic transcriptional unit. It is well

known that miRNA genes quite often form clusters in the genome (Lagos-Quintana, Rauhut et al. 2001) (Weber 2005). In our studies, we do not find any notable clusters, which could be possibly due to the in-completion of genome sequencing. Aphids have been a major pest in agriculture. Studies in molecular level for the control of aphid pest management are increasing (Akhtar, Anwar et al. 2008) (Zarrabi 2007) (Kim, Lee et al. 2008). We are very sure that miRNAs could be probably used in the pest management for controlling of insects in agricultural fields in near future. This method could make the limited use of chemical active ingredients on the crops, thereby reducing the pollution and ecological imbalances caused in recent years. Thus, we hope that this paper could be a start in understanding of miRNAs and look for their gene targets in the Aphid superfamily, *A. pisum* in particular and further their application in pest management.

References

[1] Adai, A., C. Johnson, et al. (2005). "Computational prediction of miRNAs in Arabidopsis thaliana." Genome Res **15**(1): 78-91.

[2] Akhtar, N., M. B. Anwar, et al. (2008). "Resistance to foliage feeding aphid in wheat." Pak J Biol Sci **11**(5): 801-4.

[3] Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-10.

[4] Ambros, V., B. Bartel, et al. (2003). "A uniform system for microRNA annotation." Rna **9**(3): 277-9.

[5] Arteaga-Vazquez, M., J. Caballero-Perez, et al. (2006). "A family of microRNAs present in plants and animals." Plant Cell **18**(12): 3355-69.

[6] Barakat, A., K. Wall, et al. (2007). "Large-scale identification of microRNAs from a basal eudicot (Eschscholzia californica) and conservation in flowering plants." Plant J **51**(6): 991-1003.

[7] Berezikov, E., V. Guryev, et al. (2005). "Phylogenetic shadowing and computational identification of human microRNA genes." Cell **120**(1): 21-4.

[8] Carthew, R. W. and E. J. Sontheimer (2009). "Origins and Mechanisms of miRNAs and siRNAs." Cell **136**(4): 642-55.

[9] Chan, S. P., G. Ramaswamy, et al. (2008). "Identification of specific let-7 microRNA binding complexes in Caenorhabditis elegans." Rna **14**(10): 2104-14.

[10] Chen, J., J. Olsen, et al. (2009). "A new isoform of interleukin-3 receptor {alpha} with novel differentiation activity and high affinity binding mode." J Biol Chem **284**(9): 5763-73.

[11] Cora, D., F. Di Cunto, et al. (2007). "Identification of candidate regulatory sequences in mammalian 3' UTRs by statistical analysis of oligonucleotide distributions." BMC Bioinformatics **8**: 174.

[12] Dezulian, T., M. Remmert, et al. (2006). "Identification of plant microRNA homologs." Bioinformatics **22**(3): 359-60.

[13] Fanjul-Fernandez, M., A. R. Folgueras, et al. (2009). "Matrix metalloproteinases: evolution, gene regulation and functional analysis in mouse models." Biochim Biophys Acta.

[14] Griffiths-Jones, S. (2004). "The microRNA Registry." Nucleic Acids Res **32**(Database issue): D109-11.

[15] Griffiths-Jones, S., R. J. Grocock, et al. (2006). "miRBase: microRNA sequences, targets and gene nomenclature." Nucleic Acids Res **34**(Database issue): D140-4.

[16] He, P. A., Z. Nie, et al. (2008). "Identification and characteristics of microRNAs from Bombyx mori." BMC Genomics **9**: 248.

[17] Jiang, P., H. Wu, et al. (2007). "MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features." Nucleic Acids Res **35**(Web Server issue): W339-44.

[18] Joung, J. G. and Z. Fei (2009). "Computational identification of condition-specific miRNA targets based on gene expression profiles and sequence information." BMC Bioinformatics **10 Suppl 1**: S34.

[19] Kadener, S., J. Rodriguez, et al. (2009). "Genome-wide identification of targets of the drosha-pasha/DGCR8 complex." Rna **15**(4): 537-45.

[20] Kim, H. Y., H. B. Lee, et al. (2008). "Laboratory and field evaluations of entomopathogenic Lecanicillium attenuatum CNU-23 for control of green peach aphid (Myzus persicae)." J Microbiol Biotechnol **18**(12): 1915-8.

[21] Lagos-Quintana, M., R. Rauhut, et al. (2001). "Identification of novel genes coding for small expressed RNAs." Science **294**(5543): 853-8.

[22] le Sage, C., R. Nagel, et al. (2007). "Regulation of the p27(Kip1) tumor suppressor by miR-221 and miR-222 promotes cancer cell proliferation." Embo J **26**(15): 3699-708.

[23] Li, S. C., P. Tang, et al. (2007). "Intronic microRNA: discovery and biological implications." DNA Cell Biol **26**(4): 195-207.

[24] Rodriguez, A., S. Griffiths-Jones, et al. (2004). "Identification of mammalian microRNA host genes and transcription units." Genome Res **14**(10A): 1902-10.

[25] Ruan, J., H. Chen, et al. (2008). "HuMiTar: a sequence-based method for prediction of human microRNA targets." Algorithms Mol Biol **3**: 16.

[26] Sabater-Munoz, B., F. Legeai, et al. (2006). "Large-scale gene discovery in the pea aphid Acyrthosiphon pisum (Hemiptera)." Genome Biol **7**(3): R21.

[27] Schaar, D. G., D. J. Medina, et al. (2009). "miR-320 targets transferrin receptor 1 (CD71) and inhibits cell proliferation." Exp Hematol **37**(2): 245-55.

[28] Steffen, P., B. Voss, et al. (2006). "RNAshapes: an integrated RNA analysis package based on abstract shapes." Bioinformatics **22**(4): 500-3.

[29] Thatcher, E. J., J. Bond, et al. (2008). "Genomic organization of zebrafish microRNAs." BMC Genomics **9**: 253.

[30] Watanabe, Y., M. Tomita, et al. (2007). "Computational methods for microRNA target prediction." Methods Enzymol **427**: 65-86.

[31] Weber, M. J. (2005). "New human and mouse microRNA genes found by homology search." Febs J **272**(1): 59-73.

[32] Xia, H. F., T. Z. He, et al. (2009). "MiR-125b expression affects the proliferation and apoptosis of human glioma cells by targeting Bmf." Cell Physiol Biochem **23**(4-6): 347-58.

[33] Xie, F. L., S. Q. Huang, et al. (2007). "Computational identification of novel microRNAs and targets in Brassica napus." FEBS Lett **581**(7): 1464-74.

[34] Xu, S., P. D. Witmer, et al. (2007). "MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster." J Biol Chem **282**(34): 25053-66.

[35] Yan, D., X. Zhou, et al. (2009). "MicroRNA-34a inhibits uveal melanoma cell proliferation and migration through downregulation of c-Met." Invest Ophthalmol Vis Sci **50**(4): 1559-65.

[36] Yang, Y., Y. P. Wang, et al. (2008). "MiRTif: a support vector machine-based microRNA target interaction filter." BMC Bioinformatics **9 Suppl 12**: S4.

[37] Yao, Y., Y. Zhao, et al. (2009). "Novel microRNAs (miRNAs) encoded by herpesvirus of Turkeys: evidence of miRNA evolution by duplication." J Virol **83**(13): 6969-73.

[38] Yue, J., Y. Sheng, et al. (2008). "Identification of novel homologous microRNA genes in the rhesus macaque genome." BMC Genomics **9**: 8.

[39] Zarrabi, M. (2007). "Effect of sugar beet root aphid, Pemphigus fuscicornis (Homoptera: Pemphigidae), on sugar beet yield and quality in Iran." Pak J Biol Sci **10**(19): 3462-5.

[40] Zhang, B., X. Pan, et al. (2006). "Conservation and divergence of plant microRNA genes." Plant J **46**(2): 243-59.

[41] Zhang, B., X. Pan, et al. (2008). "Identification of soybean microRNAs and their targets." Planta **229**(1): 161-82.

[42] Zhou, M., Q. Wang, et al. (2009). "In silico detection and characteristics of novel microRNA genes in the Equus caballus genome using an integrated ab initio and comparative genomic approach." Genomics **94**(2): 125-31.