Target Binding Hotspot Secrets of mir1 microRNA in *Oryza sativa* by using Bioinformatics Approaches

Mohd. Faheem Khan[1, 2], A. K. Tiwari[2], Peyush Goyal[3], A. K. Jaitly[1] and K. N. Kandpal[2]*

[1]Department of Plant Science, M. J. P. Rohilkhand University, Bareilly India;
faheemphd@gmail.com; jaitly_mjpru@yahoo.com
[2]Bioinformatics Centre, IVRI, Izatnagar, Bareilly India; aktiwari63@yahoo.com; kandpalivri@gmail.com
[3]Department of Biotechnology, Ministry of Science and Technology, New Delhi, India

Abstract

Thermodynamic properties and nucleic acid symmetry of upstream nucleotide regions to the binding sites, insertion sites or upstream to the gene regions in the eukaryotic genomes are very significant in revealing the properties of various elements such as microRNA sequence, gene sequence and Promoter sequence etc. They set signals for various microRNA to bind at specific location in the genomes and these sequences also help in gene regulation with microRNA. The miR-1 microRNA is one of the most important and common type of small microRNA precursor that regulates its target protein's expression in the cell of both plants and animals therefore its study is very significant. The main emphasis of this research and whole idea behind this kind of work is to find out upstream sequences which carry hidden properties like Railway signals and track sign board which can alert the drivers that how the way is ahead. MicroRNAs are of approximately 21-nucleotides long and play very significant role in gene regulation in variety of genes of different organisms. MicroRNAs and mRNAs (messenger RNA) constitute an important part of gene regulatory networks, influencing diverse biological phenomena. The mRNA (messengerRNA) translated in to protein by the process of translation but when microRNA binds to specific mRNA sequence it alters the gene regulation process this results the alteration in the resulting protein and its function. The binding mechanism of microRNA at target site is not very clear so the study in this field may reveal some secrets which can be helpful to understand the process of binding of microRNA at different mRNA for regulation in different genes. The analysis of upstream nucleotide sequences to the possible binding sites of microRNA in *Oryza sativa* untranslated mRNA sequences revealed some significant patterns or signals such as thermodynamic properties signals, sequence architectural signals, and energy consideration which may involve directly or indirectly for the binding of microRNA in *Oryza sativa*.

Keywords: PremicroRNA, RicemicroRNA.

1. Introduction

MicroRNA are naturally occurring, small non-coding RNA molecules, about 21–25 nucleotides in length. They play an important role in the gene expression mechanism in both animals and plants. It is also stated in the literature that rice also has good number of microRNA that takes an active part in the gene expression mechanism in several genes populated in rice genome [1–9]. We also know that how the growth and development of rice could be influenced by microRNA mediated regulation [3, 5, 33].

The main function of microRNA is to down regulate gene expression in a variety of manners, including translational repression, mRNA cleavage, and deadenylation

[10, 22, 23]. MicroRNAs play essential roles in regulating plant growth and development. In recent years, research into the mechanism and consequences of miRNA action has made great progress. With whole genome sequence available in such plants as Arabidopsis thaliana, Oryza sativa, Populus trichocarpa, Glycine max, etc., it is desirable to develop a plant miRNA database through the integration of large amounts of research information about publicly deposited miRNA data [11–13]. These data information can be very helpful to the rice research community for further research.

MicroRNAs act by binding to the complementary sites on the 3' untranslated region (UTR) of the target gene to induce cleavage with near perfect complementarity or to repress productive translation [14–19]. They were first described in 1993 by Lae and colleagues, and the term microRNA was coined in 2001 [15]. By affecting gene regulation, microRNAs are likely to be involved in most biological processes. Thousands of microRNAs since have been identified in various organisms, these microRNA complimentary to the messenger but in many case they didn't bind perfectly to the target mRNA. They also facilitate deadenylation, which leads to rapid mRNA decay [20–22]. Genetic and molecular approaches for the improvement of rice have helped establish rice as a model for plant functional genomic studies. So there should be some signals which may be responsible for controlling the binding mechanism of microRNA at target. In other words we can say that these signals may be responsible for binding the microRNA at target site.

The main emphasis of this work is to identify the most accurate microRNA targets for mir1 microRNA and understand the different signals which may be responsible for binding the mir1 microRNA at target site. This information can help us to understand the mechanism behind the binding of mir1 microRNA at target. There are several methods to predict the microRNA targets but in our work we used the support vector machine (SVM) classifier to predict the microRNA targets [7, 14–16, 24, 25]. An SVM is one of the most popular machine learning algorithm and its good performance in classification problem. Generally the efficiency and reliability of machine learning algorithm depend on choosing relevant dataset is as important as a good algorithm. A SVM builds a classifier directly from the data by investigating its characteristics. It does not require conservation information for classification. And it gave the good results for predicting the target of microRNA.

2. Material and Method

The method of experimentation was completed in the following five steps, and as described in more depth below (Figure 1):

2.1   Prediction of microRNA targets
2.2   Validation of targets
2.3   Mining of microRNA target for mir1 microRNA
2.4   Statistical analysis of to evaluate SVM prediction results
2.5   Analysis of upstream nucleotide sequences

2.1 Prediction of microRNA Targets

The below flow chart depicts the way how the novel microRNA targets prediction was done for mir1 microRNA in *Oryza sativa*. Initially the data was prepared for training and was divided into test and training sets. Then we identified different distinctive or features on which our SVM was trained [7]. Then the candidate microRNA subjected to classification. After the validation step we predicted novel targets. Experimental data were collected from the different rice databases such as TIGR and FTP web resources of NCBI etc. [31]. Untranslated mRNA sequences were also collected from these web resources. Mir1 microRNA sequence collected from already published literature [5].

2.2 Validation of Targets

Validation was done on the basis of different properties such as match, mismatch, E-value and thermodynamic properties of the sequences with the binding target sequences, Hybridization energy and accuracy of binding at both two terminals [26–28, 32].

2.3 Mining of microRNA Target for mir1 microRNA

Standalone BLAST (Basic Local Alignment Search Tool) was used for local alignment between mir1 microRNA sequence and predicted target sequences. Only those targets were taken for analysis which showed more than 75% similarity and were completely bind with both 5' end and 3' end with target sequences. Fourteen targets were predicted. Only seven targets were analyzed because these seven targets showed more significant binding properties than other.

2.4 Statistical Analysis of SVM Prediction Results for Test Set

For the evaluation of SVM results the following statistical value were used:

$$\text{Sensitivity} = \frac{\text{Number of TP}}{\text{Number of TP} + \text{Number of FN}}$$

$$\text{Specificity} = \frac{\text{Number of TN}}{\text{Number of TN} + \text{Number of FP}}$$

$$\text{Accuracy} = \frac{\text{Number of TP}}{\text{Number of TP} + \text{Number of FP}}$$
$$\frac{+ \text{Number of TN}}{+ \text{Number of FP} + \text{Number of TN}}$$

where TP (True Positive) and TN (True Negative) are correctly predicted microRNA targets sequences and target sequences, respectively similarly FP (False Positive) and FN (False Negative) are wrongly predicted non target sequence.

2.5 Extraction and Analysis of Upstream DNA Sequences

We developed BIOPERL based program that extracts hundred basepair upstream sequences to the binding site of microRNA. It was runs with the support of Standalone BLAST (Basic Local Alignment Search Tool) [29]. It also tracked the position of microRNA binding site in the Genome and checked whether the two ends i.e. 3' and 5' end of microRNA completely bind or truncated. This
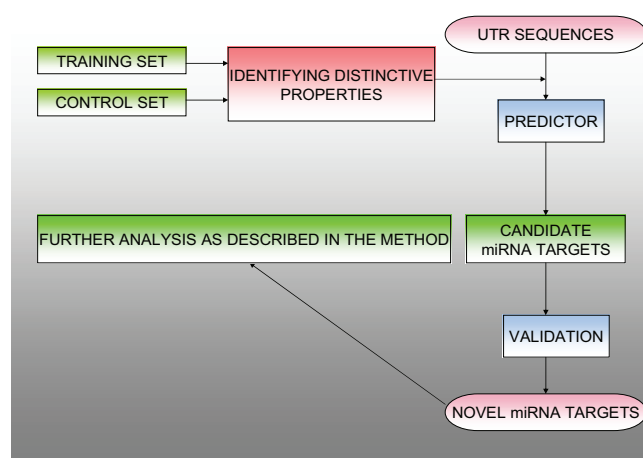

Figure 1. Flow chart for identification and the prediction of microRNA targets in Plants.

PERL based software having approximately 1000 line PERL codes it also analyzed the upstream DNA sequences and provided molecular weight of the sequences, DNA statistics i.e. the number of occurrences of each residue in the sequences, percentage of total residues, and certain groups of residues (dimmers nucleotide), This had allowed us to compare the results obtained for different upstream sequences very quickly. Graphs were developed with the help of MS-Excel and R statistical programming language [30]. Hybridization energy calculated using RNAfold for a duplex formed by the microRNA and its target. RNAfold is the part of ViennaRNA package [34], Tm value and 'GC' content was calculated by oligocalc [35].

3. Results

Fifteen novel microRNA targets were predicted by using SVM with following value of specificity (48.593), sensitivity (53.833), and accuracy (81.143) measures as indicated in Table 1.

The sensitivity and specificity are statistical measures of the performance of a binary classification test. Sensitivity measures the proportion of actual positives which are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition). Specificity measures the proportion of negatives which are correctly predicted. The accuracy of a measurement system is the degree of closeness of measurements of a quantity to its actual (true) value.

As in Table 2: 'SP' represents start position of microRNA binding site in target sequence, 'EP' represents end position of microRNA binding site in target sequence, 'HBE' represents hybridization energy between target sequence and microRNA. 'TAcc.' represents Target sequence accession number; with the help of this accession number the whole microRNA target sequence can be retrieved from NCBI web 5' and 3' are the two ends of microRNA.
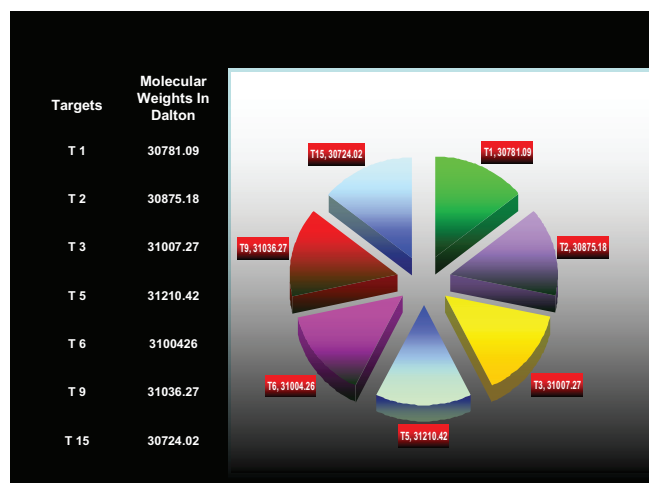
In Table 3, 'Extracted upseq' represents the upstream hundred base pairs sequences to the probable binding sites of microRNA at target sequences, 'Tm' represents melting

Table 1. Indicating the values of specificity, sensitivity and accuracy of predicted microRNA targets by SVM.

Index	Specificity	Sensitivity	Accuracy
Value	48.593	53.833	81.143

Mohd. Faheem Khan, A. K. Tiwari, Peyush Goyal A. K. Jaitly and K. N. Kandpal

5061

Table 2. Results showing the description of binding sites of mir1 microRNA.

S. No.	SP	EP	E-value	5' end	3' end	TAcc.	HBE
1.	1238	1258	9.00E-09	ok	Ok	TC249070	−25.31 k/m
2.	53	73	7.00E-10	ok	Ok	D42197	−25.31 k/m
3.	209	229	3e-009	tr	Ok	CR285364	–
4.	336	356	3.00E-09	ok	Ok	TC249077	−25.31 k/m
5.	53	73	7e-010	tr	Tr	NP654547	−25.31 k/m
6.	1175	1195	6.00E-09	ok	Ok	TC249080	−25.31 k/m
7.	328	348	2.00E-09	ok	Tr	CB680314	–
8.	4	24	2.00E-09	ok	Tr	CR284114	−32.12 k/m
9.	33	53	3.00E-09	ok	Ok	TC249081	−25.31 k/m
10.	104	124	2.00E-09	ok	Ok	CA755095	–
11.	291	311	2.00E-09	tr	Tr	TC280209	–
12.	961	981	3.00E-09	ok	Ok	NP654547	–
13.	1	21	2.00E-09	ok	Ok	BE229066	−36.06 k/m
14.	649	669	2.00E-09	tr	Tr	TC269832	–
15.	250	270	2.00E-09	ok	Ok	CA756045	−23.79 k/m

Table 3. Selected upstream sequences and their composition information

Hundred base pairs upstream sequences	Tm	%GC
CTCTGCTGTATTTAATA CAAATTTTTCCTATCTT TTTAACAGTGCTTCAGA GGGCAATGAATTCTTCT GAAGTTTTCCATTGTTA CCTGTTTTCAGTTTCA	72 ^{0}C	33
CATATCATTGGGAGTGA ATAAAGCTACGAAAGAT ACTGACTGGATAAAGCT ATATATGAAAGATACTTT GGTAAGACAGTAAACTC TTCCTGATCGTTCT	73^{0}C	35
AAGAATAATCGCAAGGCA CAGAAGTCATATCATTGG GAGTGAATAAAGCTACGA AAGATACTGACTGGATAA AGCTATATATGAAAGATAC TTTGGTAAG	73^{0}C	35
TGGATCCTTCCCATAGTAC ATGATTGTAGGAAAGCAAA TACTAGTAATCTAGGTCGAT CAAATTAAACATAATCGAAT GAAAGAACCTGTTTACAGGAAA	72^{0}C	34
CTGAAGTTTTCCATTGTTAC CTGTTTTCAGTTTCAGTAAA AAACTAAGCTCTTCAATGTA AATGGGACAACAATTTGAAT AGTTACTATAAGATCTTTTG	70^{0}C	30
TGTATTTAATACAAATTTTTC CTATCTTTTTAACAGTGCTTC AGAGGGCAATGAATTCTTCTG AAGTTTTCCATTGTTACCTGTT TTCAGTTTCAGTAAA	70^{0}C	30
GTTTTCCATTGTTACCTGTTTT CAGTTTCAGTAAAAAACTAAGC TCTTCAATGTAAATGGGACAACA ATTTGAATAGTTACTATAAGATCTTT TGGGATC	71^{0}C	31

Figure 2. Molecular weights and their pie chart distribution of upstream sequences.

temperature of microRNA this 'Tm' further indicated the stability measure of upstream target sequence essential for binding of microRNA, %GC represents the 'GC' duplex percentage in upstream sequences.

4. Discussion

Fifteen novel targets were identified for mir1 microRNA with the help of support vector machine (SVM). These fifteen targets analyzed and its hybridization energy was estimated. The start position and end position were obtained which provide the binding position of microRNA in different loci in genome. Review of e-value revealed that that each mir1 microRNA has very high quality of binding
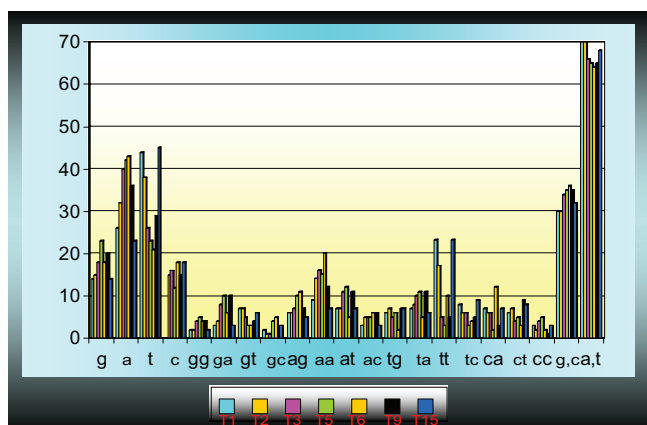
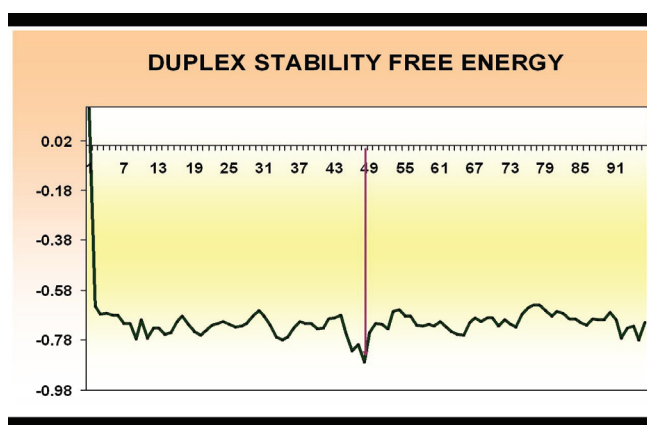**Figure 3.**   Distribution of various monomer and duplex in hundred base pair upstream sequences.

Figure 4.   Duplex stability free energy at each position of 100 base pair in selected upstream sequences.
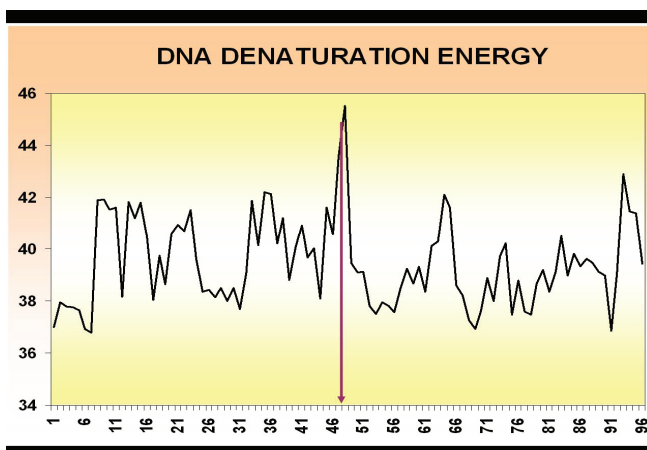

Figure 5.   DNA denaturation energy at each position of hundred base pair in selected upstream sequences.
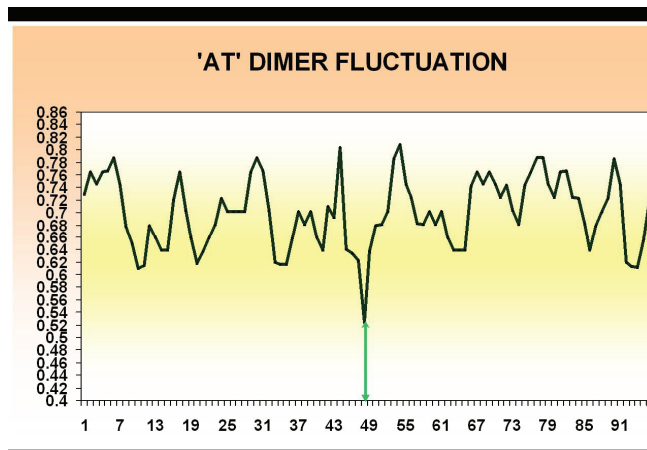

Figure 6.   'AT' duplex at each position of hundred base pair in selected upstream sequences.

behavior as each value is very small i.e. always less than zero. This further indicated that the binding behavior is very significant. In other words we can say that they are having very good binding quality.

Analysis of sequence with respect to hybridization energy (HBE) revealed that only following seven sequences have the proper identical HBE which is close to each other out of these seven sequences only one sequence has shown HBE −23.73 kcal/mol, otherwise the remaining six sequences were showing the same magnitude for HBE i.e. −25.31 kcal/mol. Targets 1, 2, 4, 5, 6, 9 and 15 successfully bind at the targets. So we took these sequences for further analysis.

The molecular weights in Dalton (Daltons is the unit to measure molecular weights) of upstream sequences obtained and presented in Figure 2. The distribution of

monomer and duplex nucleotides i.e. 'g', 'a', 't', 'c'(monomer nucleotides) and 'gg', 'ga', 'gt', 'gc', 'ag', 'aa', 'at', 'ac', 'tt', 'tc', 'ta', 'tt', 'ct', 'cc', 'gc', 'at'(duplex nucleotides) in all upstream sequences was plotted as Figure 3, and observed that the concentration of 'at' duplex was almost double or triple in comparison to other monomer and duplex nucleotides this further revealed that around binding sites of microRNA, the 'at' rich population exists as indicated in Figure 3. This further proved by Figure 6. that indicated the uniform distribution of 'AT' duplex except the position of 48 where the concentration of 'AT' drops significantly. Percentage contribution of weight of each upstream sequence were obtained and represented by pie diagram. Thermodynamic properties and 'AT' nucleotide duplex plot as in Figure 4, 5 and 6 shows common peak position in both directions

as indicated by black arrow. It may be the signals for the binding microRNA. These thermodynamic properties and 'AT' duplex plot probably can be used as a mark point for the prediction of binding site of microRNA.

We observed that all signals (molecular weights, distribution of duplexes, thermodynamic plots, 'AT' duplex plot, Hybridization energy) which almost similar or identical in all upstream sequences to the binding sites of microRNA. Thermodynamic properties were also identical in specific range i.e. it very clearly indicating the same loci by its peak i.e. 48 base pair upstream to the binding sites of microRNA as indicated and Figure 4, 5 and 6. The 'GC' content laid between 30 to 35%, melting temperature lied between 70 to 74^0C as shown in Table number.3.

We can easily concluded that some significant fluctuation happened at 48 basepair upstream sequences before the binding process of microRNA at target site and all the discussed properties (architectural, thermodynamic properties, hybridization energy, melting temperature and molecular weight) together provide the platform for the binding of microRNA. If we understands the mechanism that how microRNA selected the binding site for their function than we can also able to understand the mechanism of different gene regulation and post transcriptional modification mediated by the different microRNA from beginning to the final level. It is likely that microRNA function similar to transcription factor. Their impact on target regulation can vary from minor to significant depending on variation upstream sequences to the binding site of microRNA. So the analysis of upstream sequences to the binding site of microRNA may also help for measuring the level of different gene regulations.

5. References

1. Jones-Rhoades M W, Bartel D P et al. (2006). MicroRNAs and their regulatory roles in plants, Annual Review of Plant Biology, vol 57, 19–53.

2. Zhang B, Pan X et al. (2006). Computational identification of microRNAs and their targets, Computational Biology and Chemistry, vol 30(6), 395–407.

3. Luo Y C, Zhou H et al. (2006). Rice embryogenic calli express a unique set of microRNAs, suggesting regulatory roles of microRNAs in plant post-embryogenic development, FEBS Letters, vol 580(21), 5111–5116.

4. Doench J G, and Sharp P A (2004). Specificity of microRNA target selection in translational repression, Genes & Development, vol 18(5), 504–511.

5. Wang J F, Zhou H et al. (2004). Identification of 20 microRNAs from Oryza sativa, Nucleic Acids Res., vol 32(5), 1688–1695.

6. Khraiwesh B, Asif M A et al. (2010). Transcriptional control of gene expression by MicroRNAs, Cell, vol 140, 111–122.

7. Kim S K, Nam J W et al. (2006). microRNA target prediction using support vector machine, BMC Bioinformatics, vol 7, 407–411.

8. Schwab R, Palatnik J F et al. (2005). Specific effects of microRNAs on the plant transcriptome, Developmental Cell, vol 8(4), 517–527.

9. Pillai R S, Bhattacharyya S N et al. (2007). Repression of protein synthesis by miRNAs: how many mechanisms?, Trends in Cell Biology, vol 17(3), 118–126.

10. Rhoads R E (2010). MiRNA Regulation of the Translational Machinery, Springer Press, New York, USA.

11. Tang T, Lu J et al. (2006). miRBase: the microRNA sequence database, Methods, Methods in Molecular Biology, vol 342, 129–138.

12. Harushima Y, Nakagahra M et al. (2002). Diverse variation of reproductive barriers in three intraspecific rice crosses, Genetics, vol 160(1), 313–322.

13. Ma J, and Bennetzen J L (2004). Rapid recent growth and divergence of rice nuclear genomes, Proceedings of the National Academy Sciences, vol 101(34), 12404–12410.

14. Grün D, Wang Y L et al. (2005). MicroRNA target predictions across seven Drosophila species and comparison to mammalian targets, PLoS Computational Biology, vol 1(1), e13.

15. Lai E C (2003). microRNAs: runts of the genome assert themselves, Current Biology, vol 13(23), R925–R936.

16. Ennecke J, Stark A et al. (2005). Principles of microRNA-target recognition, PLOS Biology, vol 3(3), e85.

17. Carrington J C, and Ambros V (2003). Role of microRNAs in plant and animal development, Science, vol 301, 336–338.

18. Nelson P, Kiriakidou M et al. (2003). The microRNA world: small is mighty, Trends in Biochemical Sciences, vol 28(10), 534–540.

19. Bartel D P (2004). MicroRNAs: genomics, biogenesis, mechanism, and function, Cell, vol 116(2), 281–297.

20. Wang X J, Reyes J L et al. (2004). Prediction and identification of microRNAs and their mRNA targets, Genome Biology, vol 5(9), R65.

21. Ambros V (2004). The functions of animal microRNAs, Nature, vol 431, 350–355.

22. Giraldez A J, Mishima Y et al. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs, Science, vol 312(5770), 75–79.

23. Wu L, Fan J et al. (2006). MicroRNAs direct rapid deadenylation of mRNA, Proceedings of the National Academy Sciences, vol 103, 4034–4039.

24. Llave C, Xie Z et al. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis microRNA, Science, vol 297, 2053–2056.

25. Wang X J, Reyes J L et al. (2004). Prediction and identification of microRNAs and their mRNA targets, Genome Biology, vol 5(9), R65. Epub.

26. Sugimoto N, Nakano S et al. (2001). Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes, Nucleic Acids Res., vol. 24(22), 4501–4505.

27. Breslauer K J, Frank R et al. (1986). Predicting DNA duplex stability from the base sequence, Proceedings of the National Academy Sciences, vol 83, No. 11, 3746–3750.

28. Blake R D (1996). Encyclopedia of Molecular Biology and Molecular Medicine, Wiley-VCH, NY, vol 2, 1–19.

29. Altschul S, Gish W et al. (1990). Basic local alignment search tool, Journal of Molecular Biology, vol 215, 403–410.

30. Gentleman R (2008). R programming for bioinformatics, Chapman & Hall Press, New York, USA.

31. Ouyang S, Zhu W et al. (2007). The TIGR rice genome annotation resource: improvements and new features, Nucleic Acids Research, vol 35, 883–887.

32. Wang Y, Li Y et al. (2010). Mechanism of microRNA-target interaction: molecular dynamics simulations and thermodynamics analysis, PLOS Computational Biology, vol 6(7), 860–866.

33. Dai X, Zhuang Z et al. (2010). Computational analysis of microRNA targets in plants: current status and challenges, Brief Bioinform, vol 2, 115–121.

34. Hofacker I L (2003). Vienna RNA secondary structure server, Nucleic Acids Res., vol 31(13), 3429–3431.

35. Kibbe W A (2007). OligoCalc: an online oligonucleotide properties calculator, Nucleic Acids Research, vol 35, W43-6. Epub, 43–46.