African Crop Science Journal, Vol. 21, No. 1, pp. 75 - 83
ISSN 1021-9730/2013 $4.00

COMPUTATIONAL IDENTIFICATION OF CONSERVED MICRO RNAs FROM KODO MILLET (*Paspalum scrobiculatum*)

R. NAGESH BABU, M.N. JYOTHI, N. SHARADAMMA[1], SARIKA SAHU[2], D.V. RAI[2] and V.R. DEVARAJ[3]
Department of Biochemistry, Maharani's Science College for Women, Bangalore-560001, India
[1]Department of Biochemistry, Indian Institute of Science, Bangalore -560012, India
[2]Centre for Bioinformatics, Faculty of Biological Engineering, Shobhit University, Meerut, India
[3]Department of Biochemistry, Central College Campus, Bangalore University, Bangalore -560001, India
Corresponding author: nageshbabur@gmail.com

(Received 12 November, 2012; accepted 3 February, 2013)

ABSTRACT

Small RNA-guided gene silencing at the transcriptional and post-transcriptional levels has emerged as an important mode of gene regulation in plants and animals. Micro RNAs (miRNA) arise from long stem loop precursors which acted upon by DICER-Like enzymes. The miRNA and their precursor sequences are highly conserved among the species and this forms a key feature for their identification through homology alignment. Computational approach guides to identify the mature miRNAs as well as their precursors. The main principle behind the computational miRNA prediction is sequence and structure homologies. The *in silico* search for the homologues miRNA and their precursors among the Kodo millet ESTs enabled us to identify 4 miRNAs belonging to 3 families. A total of 34 targets were identified among which most were targeting the enzymes involved in fuel metabolism, cellular transporters, and structural proteins.

Key Words: Enzymes, *in silico*, structural proteins

RÉSUMÉ

Le petit gène silencing de l'ARN guide aux niveaux transcriptionnel and post-transcriptionnel a emergé comme un mode important de gène de regulation dans les plantes et animaux. Les Micro RNAs (miRNA) proviennent des boucles précurseurs de longues tiges qui agissent comme des enzymes DICER. Le miRNA et leurs séquences précurseurs sont hautement conservés parmi les espèces, ce qui forme une caractéristique clé pour leur identification à travers l'alignement homologue. L'approche computationnelle permet d'identifier les miRNAs en maturité ainsi que leurs précurseurs. Le principe principal dans cette prédiction computationnelle du miRNA est la séquence ainsi que la structure homologiques. L'investigation *in silico* pour les miRNA homologues et leurs précurseurs au sein du millet Kodo ESTs nous a permis d'identifier 4 miRNAs appartenant à 3 familles. Un total de 34 cibles étaient identifiés parmi lesquels étaient ciblés les enzymes impliqués dans le métabolisme énergétique, transporteurs cellulaires, et protéines structurales.

Mots Clés: Enzymes, *in silico*, protéines structurales

INTRODUCTION

MicroRNAs (miRNAs) are small RNAs that are processed from hairpin RNA precursors encoded within the genome (Bartel, 2004; Mallory and Vaucheret, 2004) and are believed to play significant roles in development within most multicellular organisms by regulating the effective level of developmentally important transcripts (Bergmann and Lane 2003; Palatnik *et al*., 2003; Chen, 2004). In plants, miRNAs tend to show greater complementarity to their targets than do animal miRNAs, and there are a number of examples of mRNA target cleavage for corresponding miRNAs in plants (Llave *et al*., 2002; Palatnik *et al*., 2003; Tang *et al*., 2003). The potential importance of miRNAs in developmental processes is evident in the mutant phenotypes associated with miRNA expression mutants (Hipfner *et al*., 2002; Brennecke *et al*., 2003; Palatnik *et al*., 2003; Chen, 2004).

There are four approaches for identifying miRNAs namely, (a) genetic screening, (b) direct cloning after isolation of small RNAs, (c) computational strategy, and (d) expressed sequence tags (ESTs) analysis. The computational miRNA prediction is based on few parameters like calculation of optimum free energies (ÄG), structural continuity, and number of G: C base pairing, etc. Direct cloning has enabled the identification of many miRNAs; however, significant variation in their expression levels has made it difficult to clone low abundance miRNAs (Lai *et al.*, 2003; Lim *et al*., 2003a).

The main drawback of all experimental methods is that they are inherently biased towards miRNAs that are highly or ubiquitously expressed and miRNAs expressed at low levels or in limited cell types may not be readily recovered. For this reason, computational approaches have been sought by a number of laboratories to complement such efforts (Grad *et al*., 2003; Lai *et al*., 2003; Lim *et al*., 2003b). Methods to date have focused on a number of different characteristics of miRNA genes, but all rely on interspecies comparative genomics at an early stage for identification. Despite the short length of miRNA sequences, the specificity of the interaction between miRNAs and their mRNA targets has demanded considerable conservation of the miRNA target sequences during evolution. Unlike animal mRNA targets, plant targets show a single sequence motif displaying a near perfect complementarity to their miRNAs. The imperfect but extensive correspondence of plant miRNAs to their mRNA targets provides a feature which makes a computational prediction more feasible. This approach is useful when data mining is performed on the basis of miRNA: mRNA targets conservation among different species (Griffith *et al*., 2003).

Kodo millet is one of the weed crops processed as an alternative to rice in drought regions of Africa and India. Owing to its anti-diabetic role, the plant gains importance as a therapeutic agent in pharmacology (Jain *et al*., 2010). Despite the limited genome resources of the millet (*Paspalum scrobiculatum*), published EST and full length nucleotide sequences in GenBank (http://www.ncbi.nlm.nih.gov/genbank) has provided the scope to get more genetic information. In this study, new miRNAs were mined in local sequence database for the purpose of understanding their roles in regulating growth and development, metabolism and other physiological processes.

METHODOLOGY

Prediction of candidate miRNA. We used EST analytical method for the identification of miRNAs. miRNAs from all known plants were downloaded from the miRNA database miRBase Release 17.16772 miRNAs are available on the site (http://www.mirbase.org/). Twelve ESTs were available on (http://www.ncbi.nlm.nih.gov). BLASTN tool was used to reveal homology between ESTs and miRNA sequences. An E-value cut-off 1.0 and word-match length 7 between query miRNAa nd ESTs sequences was used as criteria to assign identity to any sequence. The flow chart of miRNA prediction is given in Figure 1. The target sequences with no more than four mismatches were considered for secondary structure prediction using Mfold (online version). The precursor sequences were searched at 100 nucleotides upstream or downstream from the location of mature miRNAs with an increment of 10 nucleotides.

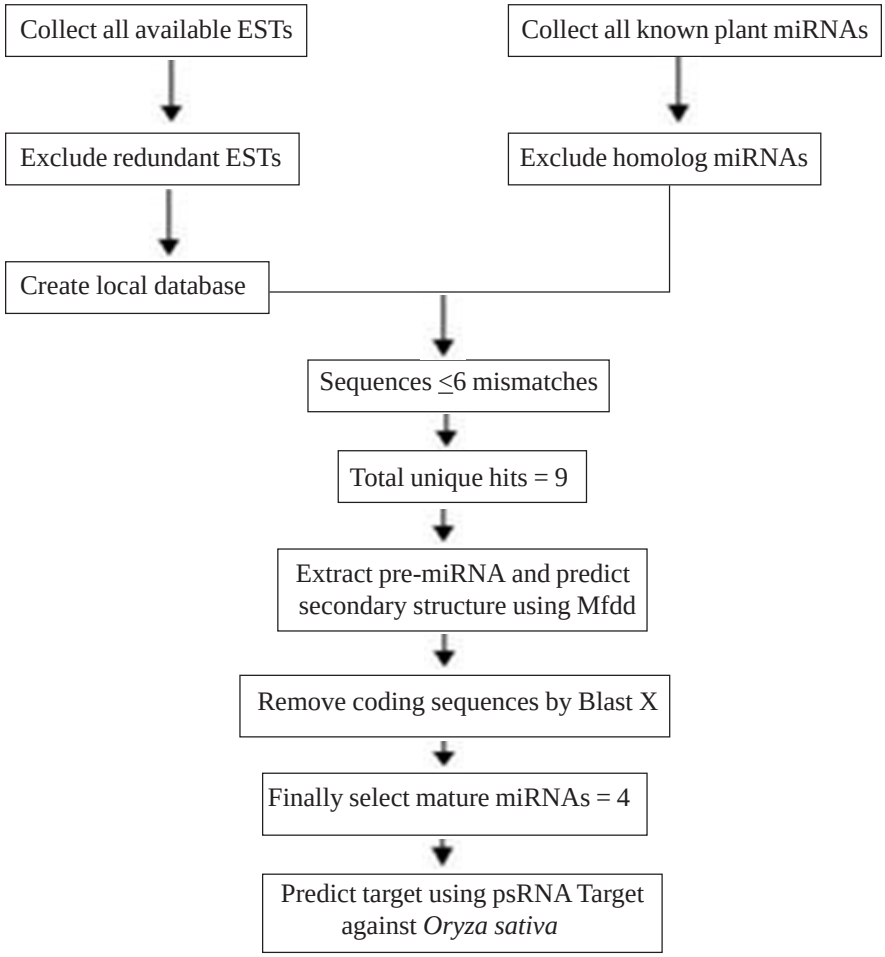

Figure 1. An overview of different steps involved in miRNA and their target prediction.

The following criteria were considered necessary for miRNA homologs (Meyers *et al*., 2008):

(i) The RNA sequence folding into an appropriate stem-loop hairpin secondary structure, (ii) a mature miRNA sequence located in one arm of the hairpin structure, (iii) miRNAs having less than 6 mismatches with the opposite miRNA sequence in the other arm, (miRNA*), (iv) no loop or break in miRNA*sequences, and (v) predicted secondary structures with higher minimal folding free energy (MFE). Also, the AU contents of pre-miRNA within 40 to 70% were considered significant.

Prediction of miRNA targets. As for Kodo millet, since only few gene sequences are available, we used Rice (*Oriza sativa)* as a reference system for finding the targets of the candidate miRNAs. The predicted tea miRNAs were used as query against the *Oriza sativa* DFCI gene index (OSGI) release 18 using miRU (http://bioinfo3.noble.org/psRNATarget/) following the criteria as (i) maximum expectation value 3; (ii) multiplicity of target sites 2; (iii) range of central mismatch for translational inhibition 9-11 nucleotide; and (iv) maximum mismatches at the complementary site ≤4 without any gaps.

Phylogenetic analysis. Due to the conserved nature of small RNAs, orthologue discovery can

be done through bioinformatics analysis. We analysed Kodo millet small RNA conservation with their orthologues. A homology search of candidate miRNAs was done against all plant miRNAs using NCBI stand-alone BLASTN, allowing a maximum of 3 mismatches and e-value <0.1. The corresponding precursor sequences of homolog small RNA's were identified and collected. The collected sequences of diverse plant species were aligned with homolog tea miRNA using Clustal W.

A query of Kodo millet small RNAs against known miRNA families (miRBase, release 15, http://www.mirbase.org/) allowed us to identify 3 previously reported large families. The precursor sequences of three known family members were selected along with respective precursor sequences of Kodo millet. Then, the maximum likely hood trees were constructed using PHYLIP, to illustrate the evolutionary relationships among the members of the family.

Nomenclature of miRNAs. The predicted miRNAs were named in accordance with miRBase. The mature sequences are designated 'miR', and the precursor hairpins are labeled as 'mir' with the prefix '*psc*'. In the cases where distinct precursor sequences have identical miRNAs with different mismatch pattern, they were named as *psc-mir-1-a* and *psc-mir*-1-*b*.

RESULTS

Prediction of putative miRNAs. A local database of Kodomillet ESTs was created with Bioedit sequence alignment editor tool (version 7.0.9.0) and was searched for putative miRNAs using the available Plant miRNAs downloaded from miRBase (http://www.mirbase.org/). Out of these, 9 sequences had less than five mismatches with previously known plant miRNAs. After carefully evaluating the hairpin structures using the criteria mentioned in the method, 4 small RNAs were finally identified from Kodo millet ESTs. The details of predicted miRNAs such as source, sequence, A+U % content and minimum folding energies were tabulated in Table 1.

The newly identified precursor miRNAs had minimum folding free energies (MFE) ranging from -103.44 to -44.25 kcal mol^{-1}, with an average of about -78.09 kcal mol^{-1} and the A + U content were ranges from 40.00 to 63.00% with an average of 55.5%. The length of the precursors ranged from 277 to 339 nt, with an average of 314 nt and mature sequences ranging from 19 to 23 nt. All the mature miRNAs were found in the stem portion of the hairpin structures (Fig. 2) containing less than 7 mismatches in the other arm, without break or loop inside the sequences. It was found that Kodo millet miRNA (*psc-miR*-2655) had been conserved with diverse plant species among monocotyledons and dicotyledons.

Phylogenetic analysis. The newly identified 4 Kodo millet miRNAs belonged to three large miRNA families (miR-9, miR398, miR2655). Two candidates of Kodo millet miRNAs were found in same family (*psc-miR*-9a and *psc-miR*-9b). The comparison of the homologues of precursor miRNA showed that most members could be having high sequence similarity and illustrates the evolutionary conservation of miRNAs among the species (Fig. 3).

Prediction of miRNA targets. A total of 34 targets was identified for 3 miRNA families based on perfect or nearly perfect complimentary pairing

TABLE 1. Predicted miRNAs from Kodo millet

miRNA	Gene ID	Mature miRNA	E- value	A+U%	MFE
psc-miR-9a	388525218	UUGCUUGGUGGUUGGGCGCU	0.039	40	103.44
psc-miR-9b	388525218	UCAGUGAAUCAUCGAAUCUUUG	0.61	63	89.05
psc-miR-398	261266876	UAGCUGCUGCUGAACAUUUGAA	0.17	59	75.62
psc-miR-2655	194302363	UGCACCUAUCCUUAUUCCUU	0.55	60	44.25

*MFE-minimal folding energies in kcal mol^{-1}

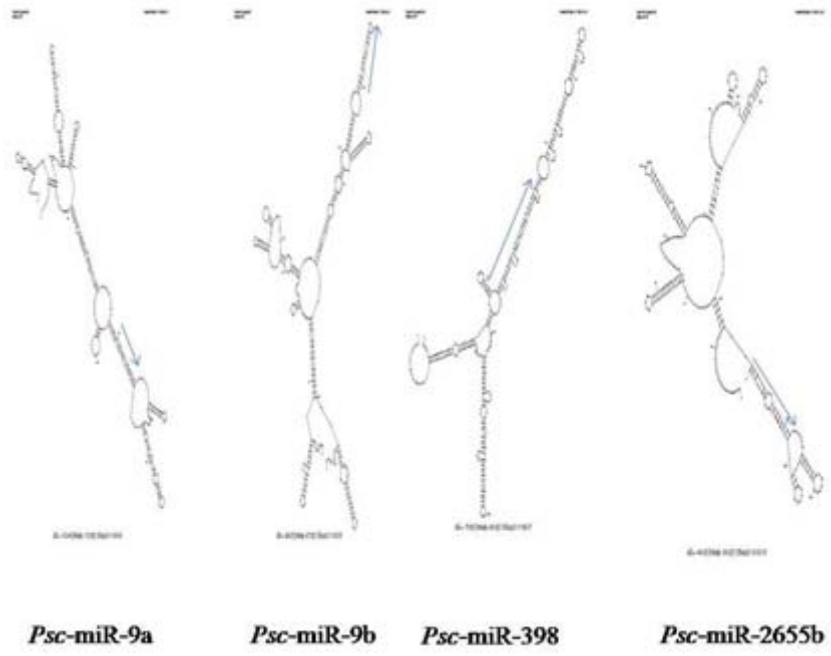

Figure 2. Predicted Stem-loop structure of Kodo millet pre- miRNAs.

between the miRNA: mRNA target. The miRNA family miR-2655 showed highest number of targets (15 targets), followed by the family miR- 9 (11 targets) and the family mir- 398 found to have least number of targets. The targets of the Kodo millet miRNAs were mostly the enzyme systems including oxygenase, hydrolase, isomerase and phosphatase. There were targets of unpredicted/ undefined proteins, which were expected to be controlling various metabolic functions. The details of miRNAs and their targets were tabulated in Table 2.

DISCUSSION

This study revealed the newly identified 4 Kodo millet miRNAs belonged to three large miRNA families (miR-9, miR398, miR2655) which form stem loop structure. These miRNA and their targets may be involved in both physiological and pathological processes make them an interesting subject of study. As to the evolutionary origin of miRNA genes, there are several different mechanisms proposed. First, miRNA genes may be generated from duplicates of protein-coding genes, or may be derived from transposable elements, or by the gene duplication of the pre-existing miRNA genes. A number of miRNA genes generated by protein -coding genes (Allen *et al*., 2004; Rajagopalan *et al*., 2006; Fahlgren *et al*., 2007; Fahlgren *et al*., 2010; Masafumi Nozawa, 2012). With constraints limited to the availability of genome data sets with Kodo millet, we were able to identify 4 miRNAs belonging to 3 large miRNA families from 12 ESTs available at the database.

Plant miRNAs came to light from *A. thaliana* studies (Sunker *et al*., 2004). Their discovery broadened the phylogenetic distribution of miRNAs to plant genomes and highlighted their ancient origin and the important role played. Recently, a significant advancement was achieved through the discovery that small RNA molecules were not only active within the cell, but also as mediators at the cell-to-cell communication level. Plant miRNAs derive from long primary transcripts (pri-miRNA) giving rise to mature RNAs products of 21-24 bp, fundamental in gene regulation. In plants, miRNAs control the degradation of messengers or restrain translation, affecting development and response to biotic and abiotic stresses. The

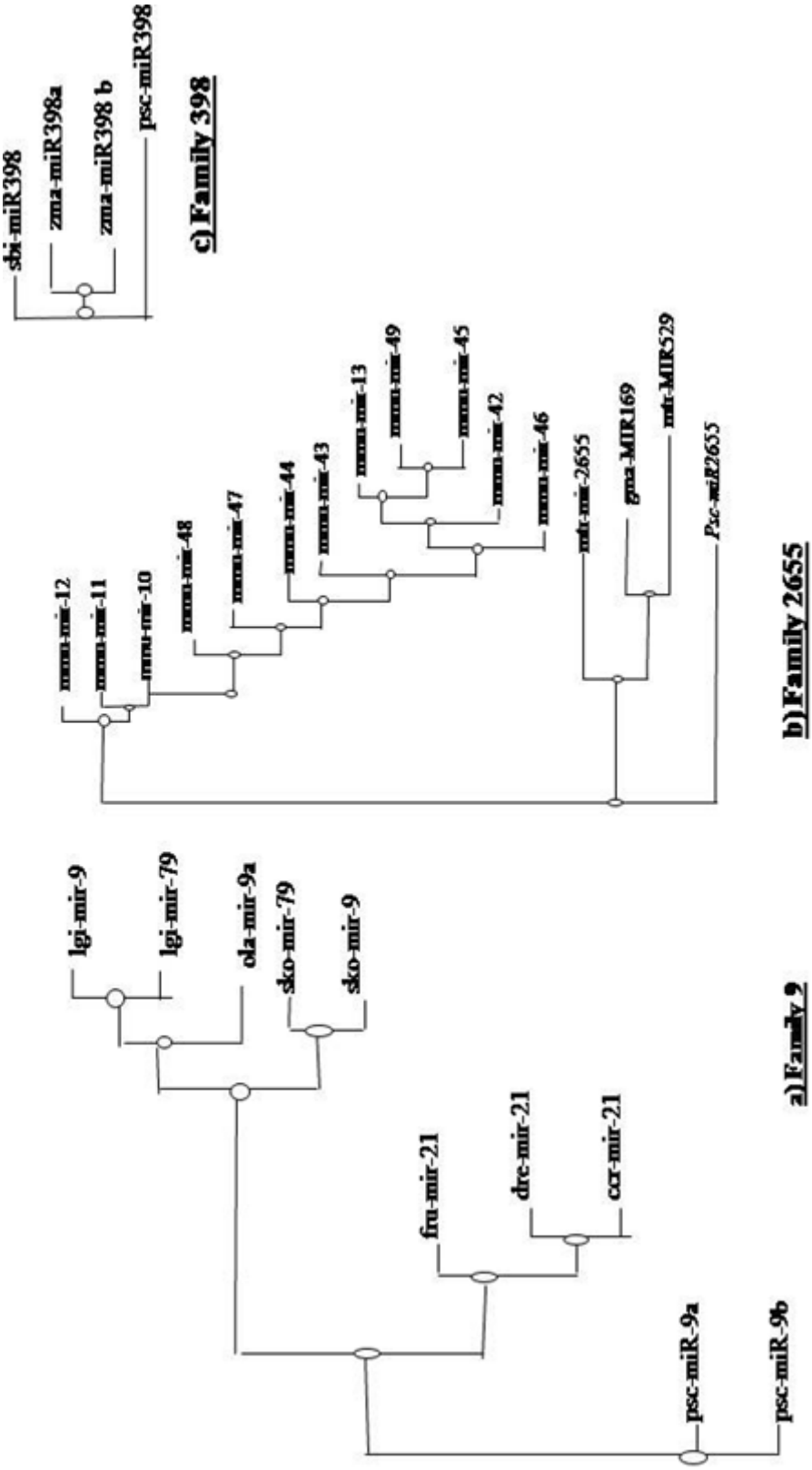

Figure 3. Phylogenetic relationships among the miRNA family members of (a) miRNA 9 (b) miRNA2655 (c) miRNA398.

TABLE 2. Targets of predicted miRNAs in Kodo millet

miRNA_Acc.	Target_Acc.	E-Value	Inhibition	Target_Description
psc-miR-9a	TC475213	2.5	Cleavage	
psc-miR-9a	TC409039	3	Cleavage	2OG-Fe(II) oxygenase
psc-miR-9a	CA756697	2.5	Cleavage	
psc-miR-9a	CR279476	2.5	Cleavage	UniRef100_Q76C86 Cluster: E3B1-2 30.8
psc-miR-9a	TC452692	2	Translation	Os03g0777600 protein
psc-miR-9a	BX899113	3	Cleavage	
psc-miR-9a	TC423506	2.5	Cleavage	DNA binding protein-like
psc-miR-9b	TC443007	2.5	Cleavage	
psc-miR-9b	TC458643	2.5	Cleavage	
psc-miR-9b	TC459837	3	Cleavage	Dual specificity protein phosphatase
psc-miR-9b	TC400961	2.5	Cleavage	Chromosome undetermined scaffold_473
psc-miR-398	CR285125	3	Cleavage	SJCHGC08647 protein
psc-miR-398	TC428192	3	Cleavage	Os06g0475800 protein
psc-miR-398	CI257505	2.5	Cleavage	
psc-miR-398	TC416281	2.5	Cleavage	Inosine/uridine-preferring nucleoside hydrolase
psc-miR-398	CI316507	3	Cleavage	
psc-miR-398	TC406923	2.5	Translation	Probable methionyl-tRNAsynthetase
psc-miR-398	TC406753	2.5	Translation	Methionyl-tRNAsynthetase, class Ia
psc-miR-398	TC470709	3	Cleavage	
psc-miR-2655	TC401656	2	Cleavage	MDR-like ABC transporter
psc-miR-2655	TC403720	2	Cleavage	Multidrug resistance protein
psc-miR-2655	CX112226	2	Cleavage	
psc-miR-2655	TC419322	2.5	Cleavage	Ankyrin repeat containing protein
psc-miR-2655	TC405151	2.5	Cleavage	TPR-like domain
psc-miR-2655	CT850662	2.5	Cleavage	Predicted protein
psc-miR-2655	EC365681	3	Cleavage	
psc-miR-2655	TC423821	3	Cleavage	C1orf167 protein
psc-miR-2655	TC447314	3	Cleavage	Acidic neurotrophin 6 beta
psc-miR-2655	TC417580	3	Cleavage	PDI(protein disulfide isomerase)-like protein
psc-miR-2655	TC483466	3	Translation	Os09g0528000 protein
psc-miR-2655	CX107827	3	Cleavage	Chromosome chr1 scaffold_5
psc-miR-2655	TC417813	3	Cleavage	Chromosome chr1 scaffold_5
psc-miR-2655	CR281776	2	Cleavage	Bone morphogenetic protein receptor,typeIA,b
psc-miR-2655	NP1105320	3	Cleavage	GB|AP004059.3|BAD21571.1 unknown protein

miRNA target gene identification is an important step for understanding the role of miRNAs in gene regulatory networks.

Although there is no striking sequence bias in the body of the mature miRNA, almost all plant miRNAs begin with a U residue (Table 1), as is seen with most animal miRNAs (Lau *et al*., 2001). Mature plant miRNAs are equally likely to be encoded in the 5'or 3'arm of the hairpin. However, when a miRNA is encoded by multiple *MIR* genes, the miRNA is always encoded in the same arm of the hairpin in all members of the gene family. This conservation in both sequence and structure implies that many of the plant miRNAs have been playing important roles since before monocots and dicots diverged approximately 250 million years ago. Because plants and animals are thought to have evolved multi-cellularity independently, and because both possess miRNAs, it appears that miRNAs have been modulating gene expression since before the emergence of multicellular life (Reinhart *et al*., 2002).

A systematic search for annotated mRNAs with complementarity to miRNAs (Rhoades *et al*., 2002; Brennecke *et al*., 2005) yielded potential

targets for most of the known plant miRNAs. The observation that most other plant miRNAs match targets with near perfect antisense complementarity led to the hypothesis that they also might act as if they were siRNAs and guide target cleavage. In our study, among the 34 targets identified, 30 were found to be inhibited by cleavage and only 4 were inhibited by translation inhibition. The pairing among the miRNA and the target mRNAs were also found to be nearly- perfect complementary with 4-6 mismatches. In several cases, particular miRNA-target mismatches have been maintained through the evolutionary distance. The identified target genes appeared to be associated with diverse biological functions. (*viz*, translation initiation factor different ion transporters MDR like ABC transporter, carbohydrate metabolism related enzymes like insulin/uridine nucleoside hydrolase, energy metabolism enzyme like Fe-oxygenase). The miRNA action by inhibiting translation is mediated through inhibition of synthesis of methionine-tRNA synthetase, the key enzyme of protein synthesis.

As the available genomic data set are limited in the case of Kodo millet, the prediction of miRNAs by computational approach is unable to give the whole set of miRNAs acting as gene regulatory elements. However, we were able to identify 4 functionally significant miRNAs belonging to 3 large miRNA families. These findings of miRNAs in Kodo millet will pave way for understanding the function and processing of small RNAs in future. Moreover, it shows a path for the prediction and analysis of miRNAs to those species whose genomes are not available through bioinformatics tools.

REFERENCES

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H. and Spatafora, J.W. 2004 Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nature Genetics* 36: 1282-1290.

Bartel, D.P. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* 116: 281-297.

Bergmann, A. and Lane, M.E. 2003. HID den targets of microRNAs for growth control. *Trends Biochemical Science* 28: 461-463.

Brennecke, J., Stark, A., Russell, R.B. and Cohen, S.M. 2005. Principles of microRNA-target recognition. *PLoS Biology*: 85-90.

Chen, X. 2004. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science* 303: 2022-2025.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J. and Sullivan, C.M. 2007. High-throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of MIRNA genes. *PLoS ONE* 2: 219-231

Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G. and Kim, J. 2003. Computational and experimental identification of *C. elegans* microRNAs. *Molecular Cell* 11: 1253-1263.

Hipfner, D.R., Weigmann, K. and Cohen, S.M. 2002. The bantam gene regulates *Drosophila* growth. *Genetics* 161: 1527-1537.

Jain, S., Bhatia, G., Barik, R., Kumar, P., Jain, A. and Dixit, V.K. 2010. Antidiabetic activity of Paspalum scrobiculatum Linn. inalloxan induced diabetic rats. *Journal of Ethnopharmacology* 127(2):325-8.

Lai, E.C., Tomancak, P., Williams, R.W. and Rubin, G.M. 2003. Computational identification of *Drosophila* microRNA genes. *Genome Biology*: 42-59.

Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B. and Bartel, D.P. 2003a. Vertebrate micro RNA genes. *Science* 299: 1540.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. 2003b. The microRNAs of *Caenorhabditis elegans Genes & Development* 7: 991 1008.

Llave, C., Xie, Z., Kasschau, K.D. and Carrington, J.C. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* 297: 2053-2056.

Mallory, A.C. and Vaucheret, H. 2004. MicroRNAs: Something important between the genes. *Current opinion of Biology 7*:120-125.

Masafumi Nozawa, Sayaka Miura and Masatoshi Nei. 2012. Origins and evolution of microRNA genes in plant species. *Genome Biology and Evolution* 4(3):230-239.

Noah Fahlgrena, Sanjuro Jogdeoa, Kristin D. Kasschaua, Christopher M. Sullivana, Elisabeth J. Chapmana,b, Sascha Laubingerc, Lisa M. Smithc, Mark Dasenkoa, Scott A. Givana, b, Detlef Weigelc and James C. Carringtona. 2010. MicroRNA Gene Evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *The Plant Cell* 22:4:1074-1089.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C. and Weigel, D. 2003. Control of leaf morphogenesis by microRNAs. *Nature* 425:257-63.

Rajagopalan, R., Vaucheret, H., Trejo, J. and Bartel, D.P. 2006. A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Development* 20: 3407-3425.

Rhoades, M., Reinhart, B., Lim, L., Burge, C., Bartel, B. and Bartel, D. 2002. Prediction of plant microRNA targets. *Cell* 110: 513-520.

Sunkar, R. and Zhu, J.K. 2004. Novel and stress-regulated MicroRNAs and other small RNAs from Arabidopsis. *Plant Cell* 16:2001-2019.

Tang, G., Reinhart, B.J., Bartel, D.P. and Zamore, P.D. 2003. A biochemical framework for RNA silencing in plants. *Genes Developement* 7:49-63.