Indian Journal of Animal Sciences **87** (1): 59–63, January 2017/Article

Prediction of novel putative miRNAs and their targets in buffalo

D C MISHRA[1], SHUCHI SMITA[2], INDRA SINGH[3], M NANDHINI DEVI[4], SANJEEV KUMAR[5], M S FAROOQI[6], K K CHATURVEDI[7] and ANIL RAI[8]

ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012 India

Received: 19 May 2016; Accepted: 21 June 2016

ABSTRACT

MicroRNAs (miRNAs) are ~22nt long non-coding RNAs, which regulate the gene regulation at the post transcriptional level in both plants and animals. These miRNA are conserved in nature and hence potential base for new miRNA prediction through homology search. No miRNAs in this species are identified so far in economically important water buffalo (*Bubalus bubalis)*. In this study, EST-based homology search, an established computational approach is used to find the potential miRNAs in buffalo. Six potential miRNA in buffalo were identified utilizing publicly available buffalo ESTs against the already known mature miRNAs of closely related species i.e. *Bos taurus*. Based on their sequence complementarity, target genes were identified which encode transcription factors (8%), enzymes (30%) and transporters (14%) as well as other proteins involved in physiological and metabolic processes (48%). These target genes also encode the proteins for signal transduction and normal development. This study will accelerate the way for further research on miRNAs and their functions in *Bubalus bubalis*.

Key words: *In-silico prediction*, miRNA, Target genes, Water buffalo

MicroRNA (miRNA) is a group of genes that do not encode proteins but regulate the expression of genes. They are 22-nucleotide-long transcripts and considered as antisense regulators of other mRNAs in most of the cases (Lee and Ambros 2001). They can bind to target messenger RNA (mRNA) transcripts of protein-coding genes and inactivate the mRNA and play an important role in gene expression by regulating a large number of target genes.

miRNAs evolve in the nucleus and transform mRNA in the cytoplasm. They reside in the introns of genomic DNA. Each miRNA is coded as approximately 300 bases in length and may either contain its own promoter region (in antisense miRNA) or become partially transcribed when the gene in which it is embedded is itself transcribed. At this stage, the miRNA is called a precursor-miRNA and acts like any other transcript. A mature miRNA is derived from its precursor miRNA called a primary RNA, and its structure can be used as a search template. A primary RNA forms a short (60-nt) stable extended stem-loop structure (or a hairpin structure) with continuous helical pairing and a few internal bulges. miRNAs are normally highly conserved in the genomes of related species, although a small number of miRNAs may be universally conserved (Bartel and Bartel 2003, Pfeffer *et al.* 2004). A miRNA has to be complementary to the 3' untranslated regions (UTRs) of a target mRNA (Bartel and Bartel 2003).

The function of an miRNA is ultimately defined by the genes it targets and the effects it has on their expression. The majority of miRNAs' target genes are computationally predicted and further verified through wet lab experiment. Unlike the situation for plants, animal miRNAs targets are difficult to predict because there exist some mismatches, gaps and G:U wobble pairs in miRNA:mRNA duplexes. Furthermore, some animal miRNAs have multiple targets on the same mRNA (Mendes *et al.* 2009). In recent past, many miRNA have been discovered in animals and plants but many other miRNA gene functions are yet to be revealed. At present, only hundreds of animal miRNA genes have been identified and associated with definite functions.

Novel miRNA is discovered in different species, using methods such as genetic screening (direct cloning through constructing a small RNA library), traditional computational approach based on whole genome sequences, and EST analysis. When the whole genome sequence of an organism is available, large number of computational approaches are used to search miRNA. EST based analysis is another method to identify miRNAs where whole genome sequences are not available. An important advantage of EST analysis is that it could provide a deeper insight in the distribution and conservation of miRNAs. For many

Present address: [1]Scientist (dcmishra@iasri.res.in), [2]Research Associate (shuchi2803@gmail.com), [3]Research Associate (indrasinghbioinfo@gmail.com). [4]M.Sc. Student (nandhumani14 @gmail.com), Centre for Plant Molecular Biology and Biotechnology, TNAU, Coimbatore, Tamil Nadu. [5]Scientist (sanjeevk@iasri.res.in), [6]Scientist (samir@iasri.res.in), [7]Senior Scientist (kkc@iasri.res.in), [8]Principal Scientist and Head (anilrai@iasri.res.in).

identified miRNAs, they are evolutionarily conserved from species to species, suggesting a powerful tool to identify miRNA homologs using the publicly available EST databases. Most of the prediction methods search for conservation of miRNA gene sequences (Bonnet *et al.* 2004) and stem-loop structures across species (Lau Nelson *et al.* 2001, Rhoades *et al.* 2002). Computational microRNA (miRNA) target prediction is one of the key means for deciphering the role of miRNAs in disease development process. Prediction of miRNA by computational means is widely accepted technique in current genomic research as it considerably save the resources.

The water buffalo (*Bubalus bubalis*) is an important animal to the lives of farmers and to the economy of many developing countries worldwide. They are good source of milk, meat, horns and skin. Evolution of large number of buffalo breeds clearly suggests the importance given to this animal as a source of food and power. Taking this into account, identification of miRNAs and its target was done for water buffalo in this study. Sequence and structure homologies were used for computer based predictions of miRNAs.

MATERIALS AND METHODS

The flow chart for the prediction of potential miRNAs in buffalo is shown in Fig. 1. The flowchart contains 3 major steps.

Identification of potential miRNA in buffalo: Buffalo ESTs were used to identify the miRNAs and their targets. For novel miRNA identification in buffalo, homology based computational method was used. ESTs were screened to select the candidate precursor miRNAs. The computational identification of mature miRNA is almost successful because the mature miRNAs are evolutionarily conserved within the different species and within their kingdom (Li *et al.* 2014). EST sequences (released March, 2015) of buffalo were retrieved from the NCBI (http://ftp.ncbi.nlm.nih.gov) and screened against the known miRNAs. These screened miRNAs were used as reference miRNA. Mature miRNA and precursor sequences from miRNA Registry Database (miRBase, Released March, 2015) were utilized to identify the potential miRNAs in buffalo. The non-redundant set of these sequences were used as query for homology search using BLASTn with number of mismatch <=4 and e-value 0.01. Resulting sequences which were considered as potential miRNA candidates only if they have at least ~ 18–20 nt length of predicted mature miRNAs and mismatch between the mature and EST was 0–4 nt. These sequences were further subjected to the BLASTx with no hits and have less than 25% similarity against NCBI nr protein (non-redundant protein) database for removing the protein coding sequences and retaining only non-coding sequences as the miRNA is a non-coding small RNAs (Altschul *et al.* 1997).

Prediction of secondary structures and putative miRNA: Secondary structure of candidate pre-miRNA is predicted using Zuker folding algorithm with MFOLD 3.6 (http://mfold.rna.albany.edu/?q=mfold/download-mfold). Overall, the precursor sequences are called as potential candidate miRNA if they follow the rules given by Meyers and others (Meyers *et al.* 2008). The criteria in brief is as follows

- The EST sequence must fold into an appropriate and proper secondary structure;
- The mature miRNA must be localized in one loop structure;

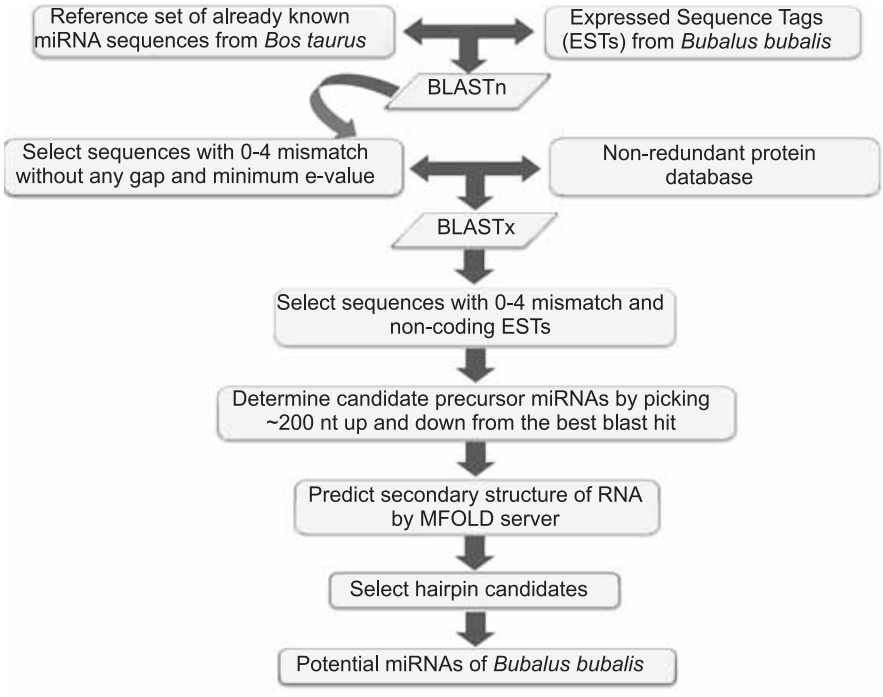

Fig. 1. Workflow of miRNA and their target prediction. Schematic representation of the buffalo miRNA search procedure used to identify homologs of known animal miRNA.

- There were no more than 6 mismatches between the predicted mature miRNA sequence and its opposite miRNA sequence in the secondary structure;
- The predicted secondary structure must contain high negative minimal folding free energy (MFE) and high minimal folding free energy index (MFEI) values which can be calculated as follows

MFEI = [(MFE/length of the RNA sequence)*100]/(G+C) %.

Predicted miRNA precursor sequences should have significantly higher MFE and MFEI.

Target identification of predicted miRNA: The general characteristic of the miRNA sequence is that it is complementary to the target gene and in some cases single miRNA is complementary to more than one target gene. Recent approaches have also shown that there is correlation between domains of miRNA expression and mRNA levels of their targets. The animal miRNA target sites are small and have only limited complementarities to their target sites. Though the miRNA degrades mRNA by different mechanism, it is important to identify the target to know the function of the newly predicted miRNA. The target prediction is accomplished by the web server based tool called TargetScanHuman 6.2 (Available at http://www.targetscan.org/) and miRanda (http://www.microrna.org) (Betel *et al.* 2008). These tools predict the biological targets of miRNAs by searching the presence of conserved 8-mers and 7-mers sites that match the seed region of each miRNA.

RESULTS AND DISCUSSION

Identification of potential miRNA in buffalo: A total of 1857 EST sequences were retrieved for miRNA prediction. Pre-processing of the ESTs was performed to maintain the quality of sequences (retained sequences with 100bp–800bp). A total of 789 sequences, were obtained. Homology search (BLASTn with < 4 mismatch; e-value 0.01) against miRBase database for potential miRNA identification was performed. After filtration with BLASTx, 238 sequences showed homology with known reference set miRNAs from mirbase.

Prediction of secondary structures: These ESTs were subjected to secondary structure prediction using MFOLD 3.6. The resulted 67 EST sequences were assumed as precursors sequences of miRNA and taken as reference set of miRNA sequences (Zhang *et al.* 2006). The MFE index (MFEI) for each sequence was calculated as previously reported by Yin *et al.* (2008). In our study, the MEFI value of predicted miRNA was from ~0.50 to 0.85. miRNA precursor sequences have significantly higher MFEI value than other non-coding or coding RNAs e.g. tRNA and rRNA. To evade false prediction of other RNAs as miRNA candidates, the MFEI is a result of structure prediction, but not considered during prediction (Meyers *et al.* 2008). The results of MFOLD were analyzed manually using Zhang *et al*. criteria (Zhang *et al.* 2006). After analysis, 6 miRNA (bbu-miR-556, bbu-miR-2350, bbu-miR-1620, bbu-miR-4062, bbu-miR-1364 and bbu-miR-325) were identified (Fig. 2). The properties of newly identified miRNAs are shown in Table 1. The comparisons of the predicted mature miRNA sequences with other members in the same family showed that most members could be found to have a high degree of sequence similarity with others. The phylogenetic tree of 2 buffalo miRNAs with the members of family illustrated the evolutionary relationships of buffalo miRNA (Fig. 3).

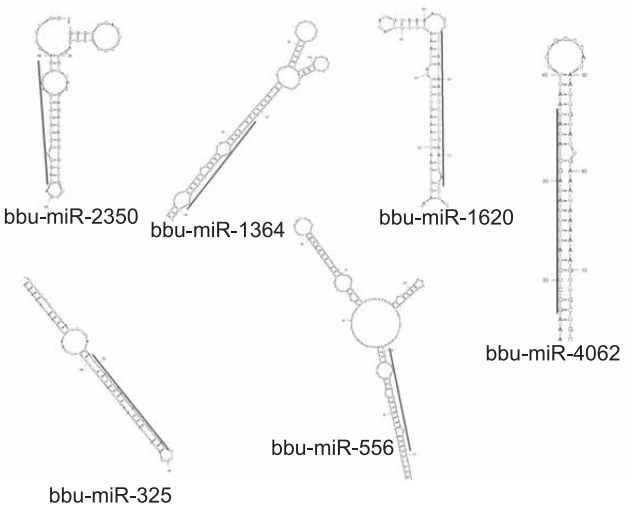

Fig. 2. Structure of the predicted miRNAs. Predicted stem-loop hairpin secondary structure of the buffalo miRNA bbu-miR-556, bbu-miR-2350, bbu-miR-1620, bbu-miR-4062, bbu-miR-1364, bbu-miR–325 identified in this study.

Potential target identification of putative miRNAs: Many targets for identified putative miRNAs were predicted that

Table 1. Properties of newly identified miRNAs in buffalo

New miRNAs	EST Id	Mature sequence	ML	LP	A+U (%)	G+C (%)	MFE	MFEI
bbu-miR-556	DT661638.1	UGAACUCUUUGAAAAGUGAG	20	160	61.25	38.75	-31.32	0.51
bbu-miR-2350	EH112563.1	CAGCUCCUGUCUCACUGC	18	120	45.83	54.16	-39.50	0.61
bbu-miR-1620	EH112616.1	UUUAAAAGCCUUUGAAUCAC	20	90	65.5	34.44	-15.50	0.50
bbu-miR-4062	GW863731.1	GCCUCUAUUUUAUAGCAU	20	90	71	28.88	-13.40	0.52
bbu-miR-1364	HO000109.1	CAAAGUUGGAAGGCUGGA	18	150	42.66	37.33	-45.63	0.82
bbu-miR-325	HO004831.1	UUUAUUGAGCAUCUCUUAU	19	122	68	31.96	-23.10	0.60

ML, length of mature miRNAs; LP, length of precursor; A+U, (adenine+uracil)%; MFEI, minimal folding free energy index.

Table 2. List of the potential targets of newly identified miRNA bbu-miR-556 and bbu-miR-2350 in buffalo

miRNA	Gene name	Biological function of gene
bbu-miR-556	zinc finger protein 215	• DNA binding • Metal ion binding • Sequence specific DNA binding RNA polymerase II transcription factor activity • Sequence-specific DNA binding transcription factor activity
	Mediator complex subunit 12-like	• RNA polymerase II transcription cofactor activity • Beta-catenin binding • Transcription factor binding
	DnaJ (Hsp40) homolog, subfamily C, member 19	• Protein folding • Protein targeting to mitochondrion • Visual perception
	Peptidylglycine alpha-amidating monooxygenase	• Ca, Cu, Zn ion, L-ascorbic acid binding • Protein and protein kinase binding
	calpain 6	• Calcium dependent cysteine–type endopeptidase activity • Microtubule binding
	Myotrophin	• Cell growth, neuron differentiation • Regulation of translation • Skeletal muscle cell differentiation
bbu-miR-2350	Adaptor-related protein complex 1, sigma 1 subunit	• Protein transporter activity
	TNF receptor-associated factor 6	• Ligase activity, protein binding, protein kinase and kinase B binding, thioesterase binding, zinc ion binding
	Meis homeobox 1	• Protein and sequence specific DNA binding
	Cytohesin 3	• ARF guanyl-nucleotide exchange factor activity • Phosphatidylinositol-3,4,5-trisphosphate binding

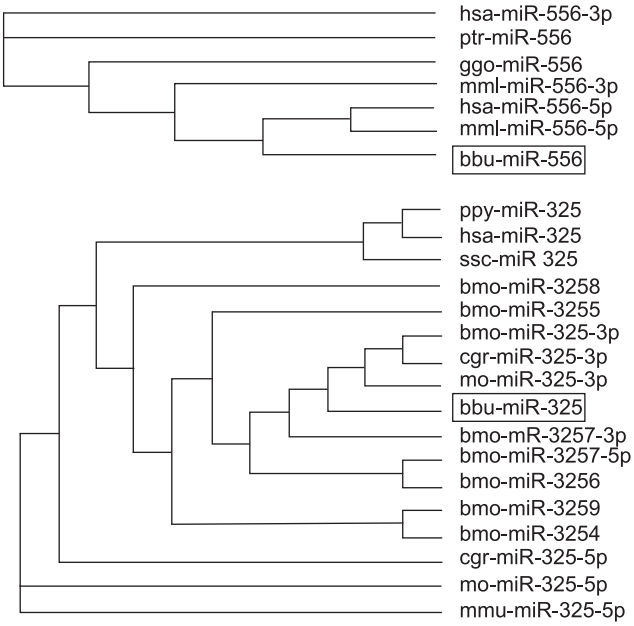

Fig. 3. Phylogenetic tree of the predicted miRNAs. Phylogenetic tree for the newly identified miRNAs in buffalo (bbu-miR) showing homology with the several published miRNAs of miR556 and miR325 family.

belong to several gene families with different biological functions. Our prediction of target genes for the buffalo miRNAs discovered that more than one gene was regulated by individual miRNA. These finding also suggested that miRNA research should be focused on regulatory networks rather than individual connections between miRNA and their predicted target genes or regulatory factors. The bbu-miR-556 and bbu-miR-2350 miRNAs target protein binding, which directly or indirectly affect growth and development, and also specific genes which control cell growth and neuron differentiation in animals. Genes targeted by our identified miRNAs contain binding activity. The miRNA target genes that bind with DNA are Zn, Cu, Ca, L-ascorbic acid, kinase protein, microtubule and Kinase B. The list of bbu-miR-556 and bbu-miR-2350 targets are shown in detailed manner in Table 2. The miRNA bbu-miR-1620 targets sex determining region–Y box 6 (SOX-6) which is conserved in chimpanzee, dog, cow, mouse, rat, chicken, zebra fish and frog. This gene involved in the process of astrocyte differentiation, cardiocyte differentiation, cartilage development, cell morphogenesis, cellular response to transforming growth factor beta stimulus, erythrocyte development, gene silencing, in utero embryonic development, muscle cell differentiation, muscle organ development, negative regulation of transcription from RNA polymerase II promoter, oligodendrocyte cell fate specification, positive regulation of cartilage development, chondrocyte differentiation and mesenchymal stem cell differentiation, post-embryonic development. It also affects the gene encoding the transient receptor potential cation channel, subfamily C, member 4 associated proteins which involved in the phosphate and ion binding. This also binds with vascular endothelial growth factor A. This gene is a member of the PDGF/VEGF growth factor family and encodes a protein that is often found as a

disulphide linked homodimer. This protein is a glycosylated mitogen that specifically acts on endothelial cells and has various effects, including mediating increased vascular permeability, inducing angiogenesis, vasculogenesis and endothelial cell growth, promoting cell migration, and inhibiting apoptosis. Elevated levels of this protein are linked to POEMS syndrome, also known as Crow-Fukase syndrome. Mutations in this gene are associated with proliferative and non-proliferative diabetic retinopathy. Alternatively spliced transcript variants, encoding either freely secreted or cell-associated isoforms, have been characterized. There is also evidence for the use of non-AUG (CUG) translation initiation sites upstream of, and in-frame with the first AUG, leading to additional isoforms.

Overall, these findings suggested that the buffalo miRNAs may target both transcription factors as well as some specific protein binding genes. These findings may help in understanding the function and processing of small RNAs of buffalo in future.

In this study, 6 putative and novel miRNAs were identified in buffalo through *in-silico* approach based on EST data available in public domain. In addition to this, the target genes and their functions of these miRNA were also predicted. Our results will further help to understand the gene regulation mechanism and the biogenesis of miRNA. This research can be extended through developing regulatory networks between miRNAs and their targets through system biology approach.

ACKNOWLEDGEMENT

We acknowledge CABin Scheme of ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi for providing all the required facility to conduct this study.

REFERENCES

Altschul S F , Madden T L, Schäffer A A, Zhang J, Zhang Z, Miller W, Lipman D J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids Research* **25**: 3389–3402.

Bartel B and Bartel D P. 2003. MicroRNAs: at the root of plant development? *Plant Physiology* **132**: 709–17.

Betel D M, Wilson A, Gabow A, Marks D S and Sander C. 2008. The microRNA.org resource: targets and expression. *Nucleic Acids Research* **36**: D149-D153.

Bonnet E J, Wuyts P, Rouzé and Van de Peer Y. 2004. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* **20**: 2911–17.

Lau Nelson C L, Lim E Weinstein and Bartel D. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858–62.

Lee R C and Ambros V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**: 862–64.

Li X *et al*. 2014. Computational identification of conserved microRNAs and their targets from expression sequence tags of blueberry (*Vaccinium corybosum*). *Plant Signaling and Behavior* **9**: e29462.

Mendes N D, Freitas A T and Sagot M F. 2009. Current tools for the identification of miRNA genes and their targets. *Nucleic Acids Research* **37**: 2419–33.

Meyers B C *et al*. 2008. Criteria for annotation of plant MicroRNAs. *Plant Cell* **20**: 3186–90.

Pfeffer S *et al.* 2004. Identification of virus-encoded microRNAs. *Science* **304**: 734–36.

Rhoades M W *et al.* 2002. Prediction of plant microRNA targets. *Cell* **110**: 513–20.

Yin Z C, Li X, Han and Shen F. 2008. Identification of conserved microRNAs and their target genes in tomato (*Lycopersicon esculentum*). *Gene* **414**: 60–66.

Zhang B, Pan X, Cannon C H, Cobb G P and Anderson T A. 2006. Conservation and divergence of plant microRNA genes. *Plant Journal* **46**: 243–59.