Journal of Cell and Molecular Research (2012-2013) 4 (2), 62-67

Computational Analyses for Identification Novel MicroRNAs from Cattle and Sheep

Balal Sadeghi[1], Mohammadreza Nassiri[1*], Ali Masoudi-Nejad[2], Mojtaba Tahmoorespour[1], Hesam Dehghani[3] and Hamed Ahmadi[4]

[1]*Department of Animal Science, and Research Institute of Biotechnology, Ferdowsi University of Mashhad, Mashhad, Iran*
[2]*Laboratory of System Biology and Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*
[3]*Embryonic and Stem Cell Biology and Biotechnology Research Group, Research Institute of Biotechnology, and Department of Basic Science, Faculty of Veterinary Medicine, Ferdowsi University of Mashhad, Mashhad, Iran*
[4]*School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran*

Received 25 November 2012 Accepted 13 February 2013

Abstract

MicroRNAs (miRNA) are a class of noncoding and regulatory RNA molecules about 22 nucleotides in length. MicroRNAs regulate gene expression by an RNA interfering pathway through cleavage or inhibition of the translation of target mRNA. Many miRNAs have been reported for their important roles in developmental processes in various animals, but there is limited information about cattle and sheep miRNAs. The comparative genomics approach due to their conserved nature is a good source for the miRNAs discovery. Cattle and sheep are ideal model organisms for biological and comparative genomics studies. In our study, a computational method based on expressed sequence tag (EST) analysis was used for detection of cattle and sheep miRNAs. In cattle, 25 miRNA candidates found by homology searching frequently clustered at certain chromosomes and 28 miRNAs in sheep had been detected. Our results show that the cattle and sheep miRNA database can be providing useful information for investigating biological functions of miRNAs in cattle and sheep. Furthermore, the bioinformatics approach is a good manner for studying these functions.

***Keywords*:** microRNAs, comparative genomics, homology searching, sheep, cattle

Introduction

MicroRNAs (miRNAs) are non-coding RNAs, which are approximately 22 nucleotides (nt), and can regulate the expression of target genes by binding to complementary sites. It is more prevalent for miRNAs to down-regulate the expression of target genes by binding to the complementary sites in transcripts and cause transcript degradation or translational repression (Bartel, 2004; Pillai et al., 2007). However, recent studies found that miRNAs could increase protein translation by binding to the complementary promoter sequences (Place et al., 2008; Vasudevan et al. 2007).

The miRNAs in different organisms are involved in many gene regulation processes like; growth and development, transformed genes inactivation, cell signaling pathways, external stresses, cancer related proteins and defense against the invading viruses (Kidner and Martienssen, 2005).

The first miRNA was discovered in larval mutants of the nematode *Caenorhabditis elegans* and was identified as a down-regulator of gene expression. Larval worms with mutations in the lin-4 gene showed defects in the timing of cell division, and the miRNA encoded by the lin-4 gene silenced the expression of lin-14 mRNA (Lee et al., 1993). Hundreds of miRNAs have been identified by small RNA cloning and computational analysis in plants and animals (Bentwich et al., 2005). The view of miRNA biogenesis and maturation holds that compartmentalized stepwise processing of miRNAs takes place first in the nucleus and then in the cytoplasm. The primary transcripts of miRNAs (pri-miRNAs) are processed in the nucleus by the RNase III enzyme, Drosha, partnered by DGCR8 (or Pasha in *C. elegans*), to stem-loop intermediates of about 70 nt known as pre-miRNAs. The pre-miRNAs are then transported by Exportin-5 to the cytoplasm for cleavage by Dicer (another RNase III enzyme) and maturation to their active forms, which are taken in by an RNA interfering silencing complex. The miRNAs drive their target mRNAs to be cut or translationally inhibited, depending on the complementarity of the miRNA sequences to their targets (Plasterk, 2006).

There are basically two kinds of approaches to identify miRNAs. One is to sequence size-

∗Corresponding author E-mail: nassiryr@um.ac.ir

fractionated cDNA libraries. Many known miRNAs have been identified by this method (Fu et al., 2005). Mentioned method allows the identification of both conserved and unconserved miRNAs, but a limitation of which is that some miRNAs are expressed at low levels, and expressed at very specific stages or in rare cell types. In contrast, computational strategies, which may especially miss those that are not phylogenetically conserved, provide an efficient way to predict miRNAs and their targets by surveying genomic sequences or other databases like expressed sequence tags (ESTs). Which are based on the secondary structure characteristics, phylogenetic conservation of both sequence and structure, and thermodynamic stability of hairpins. Computational approaches have been successfully applied in vertebrates (Huang et al., 2008), insect (Singh and Nagaraju, 2008), and plant (Jin et al., 2008).

Cattle and sheep are two of the most important agricultural livestock for meat production, but there is limited information about their miRNAs (miRBase 18.0, November 2011) (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006; Griffiths-Jones et al., 2008). The aim of this study was identification more miRNAs in cattle and sheep in order to increase our knowledge and understanding of the gene regulatory networks in these livestock.

Materials and Methods

Computational identification of cattle miRNAs

In order to identify potential cattle miRNAs, we downloaded all of the known mammalian (human, mouse, pig, sheep, and dog) miRNAs registered in the database of miRBase 18.0 (http://www.mirbase.org/) (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006; Griffiths-Jones et al., 2008) and regarded them as starter reference sequences. Cattle miRNAs registered in the database were used to test the prediction of miRNA candidates. Cattle genomic survey sequences, mRNA, cDNA, RefSeq protein database, and EST sequences were obtained from NIH Genbank nucleotide databases (http://www.ncbi.nlm.nih.gov/). After removal of repeat miRNAs obtained from the database of miRBase, the rest of the unique miRNAs were blasted against the cattle genome (BTAU4.0). Sequences encoding protein were ignored by blasting the RefSeq protein database, and the secondary structures of the rest of the sequences flanked with 50–60 nt of genomic sequence were predicted by the software Mfold (version 3.2) (Zuker, 2003), at 37° C and 1 M NaCl. The sequences were considered to be miRNA candidates if they met the criteria as previously described by Zhang et al. (2006), with some modifications according to the features of most of the known miRNA hairpin precursors: (1) a sequence can fold into an appropriate stem-loop hairpin; (2) there is a mature miRNA sequence site in one arm of the hairpin structure; (3) there are predicted stem-loops without large loops or bulges in the miRNA* sequences; (4) predicted mature miRNAs have no more than four nucleotide substitutions compared with known miRNAs; and (5) predicted secondary structures have a minimum free energy (dG) higher than -18.8° C and paired bases within the stem-loop hairpin ≥16.

Computational identification of sheep miRNAs

Animal miRNAs and their precursor sequences were obtained from miRBase (Release 18:0 November 2011 at http://www.mirbase.org/) (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006; Griffiths-Jones et al., 2008). The Sheep ESTs (total 338483 sequences) were downloaded from NCBI database (http://www.ncbi.nlm.nih.gov/). A five-step prediction method was used to identify Sheep miRNAs. First, alignment of known animal miRNAs was conducted by BioEdit software package (Hall, 1999) to remove redundant sequences. Second, we used remaining miRNAs as query sequences for BLAST searched against the Sheep ESTs with BLASTN, the parameter settings were as follows: E-value cut-off was 10, the number of descriptions and alignments were 1000. Third, the ESTs with no more than two mismatched nucleotides were chosen to blast with the known pre-miRNA sequences. Fourth, the ESTs which have >90% similarity with the corresponding known precursor sequences were selected to remove the repeated sequences and the protein-coding sequences by BLASTN and BLASTX program. The last step was to apply Mfold (version 3.2) (Zuker, 2003), to further identify the pre-miRNAs. Four criteria were used: (1) the A+U content of the precursor sequences should range from 30 to 70%; (2) the mature miRNA should locate on one arm of the hairpin structure, and the corresponding positions of the mature miRNAs in their pre-miRNAs were nearly identical, which was calculated by δ –len (A,B) [13]. We used δ -len cut-off 10 as the default parameter; (3) the minimum free energy (MFE) of the secondary structure for each potential pre-miRNA was less than -20 kcal/mol; (4) the hairpin must include at least 16 bp within the first 22 nt of the miRNA, and should not contain large internal loops or bulges, particularly not large asymmetric bulges, as described in

reference (Clop et al., 2006). The Sheep miRNA, oar-mir-134, was selected for conservation and phylogenetic analyses. The analyses of oar-mir-134 with human (*H. sapiens*), mouse (*M. musculus*), rat (*R. norvegicus*) and cattle (*B. Taurus*) orthologues were done by the publically available weblogo: a sequence logo generator (Crooks et al., 2004) and ClustalW (Larkin et al., 2007) to generate cladogram tree using neighbor joining clustering method respectively.

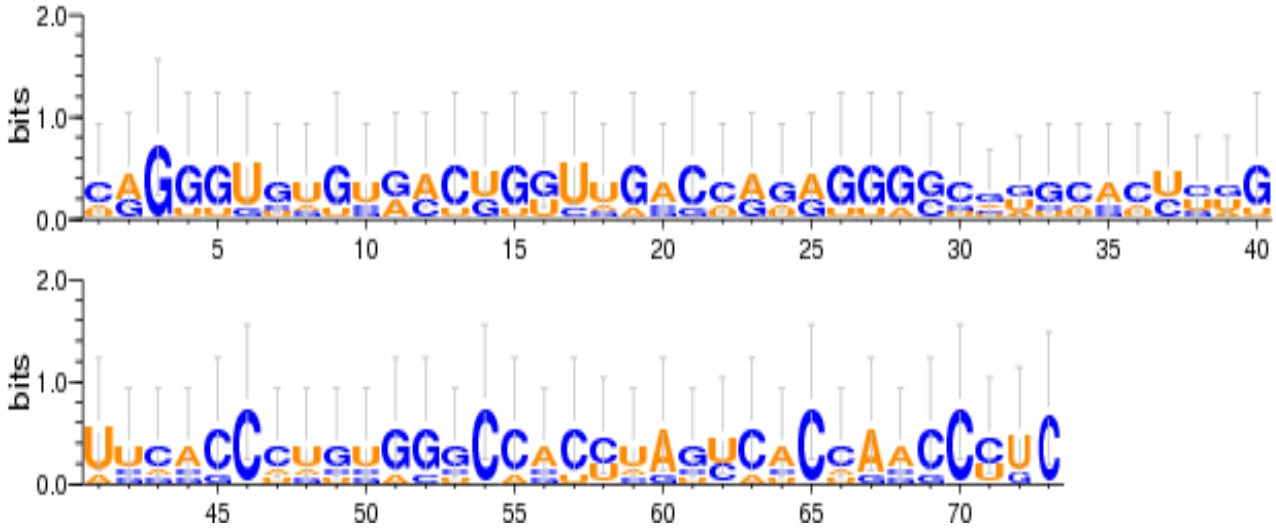

Figure 1. The miRNAs conservation studies. Alignment of sheep (*O. aries*) pre-miRNAs with human (*H. sapiens*), mouse (*M. musculus*), rat (*R. norvegicus*) and cattle (*B. Taurus*) was generated using Weblogo.

Results

Identification of 25 cattle miRNA candidates by computational prediction in order to identify potential cattle miRNAs by homology searching was carried out. We downloaded all of the miRNAs for humans, mice, pigs, dogs, and sheep registered in the miRBase (http://www.mirbase.org/). After removal of repeated miRNAs, 251 miRNAs were considered as starter references to blast the cattle genome. Of the 11121 sequences targeted within the 0–4 mismatch, only 1052 had a match of at least 16 nt, with >80% homology to known miRNAs, and were within two mismatches of a known miRNA of less than 19 nt. Of those sequences, 94 were fragmented sequences of encoding protein or of known noncoding RNA. The secondary structure of the remaining sequences was predicted by the program Mfold 3.2, and only 81 hairpins met the criteria for miRNAs. After blasting against the cattle genome EST database, we finally identified 25 cattle miRNA candidates. The cattle miRNA candidates were designated with the matched known miRNAs. The result showed that many of the candidates were likely expressed from two or three miRNA genes located on different chromosomes, such as bta-let-7a, bta-let-7f, bta-mir-1 and bta-mir-124a.

The homology search through comparative genetics is a rational approach to find interesting findings. The homology based search and applying in silico approach resulted 28 miRNAs in sheep. These 28 miRNAs are predicted from the analyses of 338483 sheep ESTs. First, we used a computational approach to exploit Sheep miRNAs in sheep EST database. Following a set of strict filtering criteria, we finally identified 28 Sheep miRNAs.

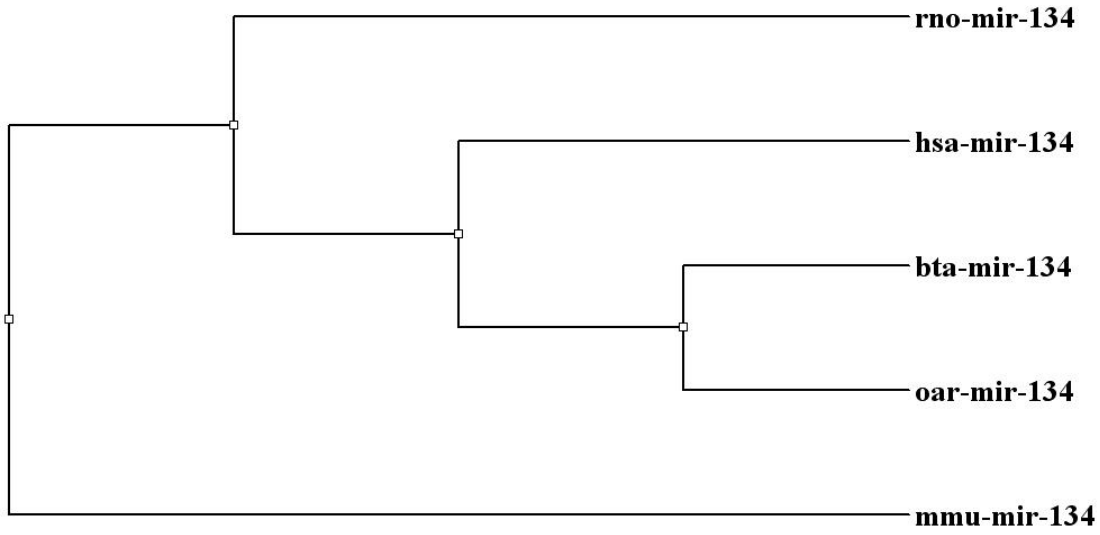

Figure 2. The sheep miRNAs phylogenetic analysis. The Phylogenetic analysis of the sheep pre-miRNAs with human (*H. sapiens*), mouse (*M. musculus*), rat (*R. norvegicus*) and cattle (*B. Taurus*) was done with the help of Clustalw and cladogram tree was generated using neighbor joining clustering method. The Phylogenetic tree showed that sheep is more closed to *B. Taurus* (bta).

Discussion

MicroRNAs represent a large class of gene regulatory molecules that control fundamental cellular processes in animals, including control of developmental timing, hematopoietic cell differentiation, apoptosis, cell proliferation, and organ development (Bartel, 2004; Ambros, 2004). miRNAs and their targets seem to form complex regulatory networks (Lewis et al., 2005). Identification of miRNAs would benefit biologists by improving our understanding of the miRNA regulatory networks. Small RNA cloning and genetic manipulations were tackled to discover novel miRNAs, and computational algorithms were also successfully applied to identify hundreds of miRNAs (Grad et al., 2003; Lagos-Quintana et al., 2002; Lai et al., 2003; Ruby et al., 2006). Computational prediction would overcome the difficulty of biochemically discovering low-abundance miRNAs.

In this study, we identified 25 miRNAs in cattle by homology searching. Coutinho et al., (2007) reported 129 mature cattle miRNAs identified by homology searching and small RNA cloning from immune-related tissues. They started the search, however, only from human miRNA stem-loop sequences or combined the predicted human miRNAs. Consequently, their prediction ignored other miRNAs not found in humans but in other mammals, such as mice, pigs, sheep, and dogs. Furthermore, their prediction was based on the draft genome of cattle, at almost the same time; Gu et al. (2007) discovered 59 distinct miRNAs from cattle adipose and mammary gland tissues by small RNA cloning. Combining the two reports, about 130 miRNAs have been discovered from adipose, mammary gland, and immune-related tissues in cattle. To identify more miRNAs in cattle, Long et al. (2009) searched experimental miRNAs based on the cattle genome. They also cloned miRNAs from four tissues of brain, liver, lung, and heart, given that miRNAs were probably expressed with tissue specificity. Our study results covered all miRNAs identified in previous studies (Coutinho et al., 2007; Gu et al., 2007; Long and Chen, 2009). This indicated that cattle miRNA candidates were predicted at very high accuracy.

We finally identified 28 Sheep miRNAs. Among which, 13 identified miRNAs sequences were identical or highly similar (1 or 2 nt mismatches) with those of known miRNAs in related species, which agreed with that most miRNAs are conserved among mammalian species. The minimum free energy (MFE) of all precursor sequences was lower than -20 kcal/mol. However, the MFE of its homologs in other species also has a higher value, for example, -16.7 kcal/mol in *Homo sapiens* and -15.2 kcal/mol in *Pan troglodytes* (Ambros et al., 2003; Lagos-Quintana et al., 2001).

The identified sheep's miRNA secondary structures showed that there are at least 16 nt engaged in Watson–Crick or G/U base pairings between the mature miRNA and the opposite arms (miRNAs*) in the stem region except few, where the reference miRNAs have also less base pairings and the hairpin precursors do not contain large

internal loops or bulges. Similar findings were reported by many groups (Barozai et al., 2008; Barozai et al., 2012; Glazov et al., 2009; Hossain et al., 2009; Barozai et al., 2011; Barozai et al., 2011).

The sheep pre-miRNA (oar-mir-134) conservation and phylogenetic analyses with human (*H. sapiens*), mouse (*M. musculus*), rat (*R. norvegicus*) and cattle (*B. Taurus*) revealed that approximately miRNA (miR 134) sequences showed conserved nature in these animals (Fig. 1). The Phylogenetic analysis of the same miRNA (mir-134) sequences suggested that the sheep is more closed to *B. Taurus* (Fig. 2). The results are in agreement with the previous reported works (Lai et al., 2003; Sheng et al., 2010; Huang et al., 2010).

In conclusion, this study identified the 25 and 28 miRNAs in cattle and sheep, respectively. The miRNAs are the good functional genomic resources to understand the gene regulatory mechanism in cattle and sheep. These cattle and sheep's miRNAs will be useful in the near future for the improvement of meat production from this most important livestock animal. The identification of miRNA targets and expression profiles at a genome wide level, perhaps help us to understand why miRNAs are clustered at a genomic locus, too.

Acknowledgment

The research for this paper was financially supported by the Iran National Science Foundation, Deputy of Science and Technology, Presidential Office, grant no. 90007420.

References

1- Ambros V. (2004) The functions of animal microRNAs. Nature 431: 350–355.

2- Ambros V., Bartel B., Bartel D. P., Burge C. B., Carrington J. C., Chen X., Dreyfuss G., Eddy S. R., Griffiths-Jones S., Marshall M., Matzke M, Ruvkun G, Tuschl T. (2003) A uniform system for microRNA annotation. RNA 9: 277–279.

3- Barozai M. Y., Baloch I. A. and Din M. (2011) Computational identification of MicrorNAS and their targets in two species of evergreen Spruce tree (*Picea*). World Academy of Science, Engineering and Technology 75: 413–418.

4- Barozai M. Y., Din M. and Baloch I. (2011) Identification of microRNAs in ecological model plant Mimulus. Journal of Biophysical Chemistry 2: 322–331.

5- Barozai M. Y., Baloch I. A. and Din M. (2012) Identification of MicroRNAs and their targets in Helianthus. Molecular Biology Reports 39: 2523-32.

6- Barozai M. Y., Irfan M., Yousaf R., Ali I., Qaisar U., Maqbool A., Zahoor M., Rashid B., Hussnain T. and Riazuddin S. (2008) Identification of micro-RNAs in cotton. Plant Physiology and Biochemistry 46: 739–751.

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

8- Bentwich I., Avniel A., Karov Y., Aharonov R., Gilad S., Barad O., Barzilai A., Einat P., Einav U., Meiri E., Sharon E, Spector Y, Bentwich Z. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nature Genetics 37: 766–770.

9- Clop A., Marcq F., Takeda H., Pirottin D., Tordoir X., Bibé B., Bouix J., Caiment F., Elsen J. M., Eychenne F., Larzul C, Laville E, Meish F,Milenkovic D, Tobin J, Charlier C, Georges M. (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nature Genetics 38: 813–818.

10- Coutinho L. L., Matukumalli L. K., Sonstegard T. S., Tassell C. P. V, Gasbarre L. C., Capuco A. V., Crooks G. E., Hon G., Chandonia J. M. and Brenner S. E. (2004) Web- Logo: a sequence logo generator. Genome Research 14: 1188–1190.

11- Fu H., Tie Y., Xu C., Zhang Z., Zhu J., Shi Y., Jiang H., Sun Z. and Zheng X. (2005) Identification of human fetal liver miRNAs by a novel method. FEBS Letters 579: 3849–3854.

12- Glazov E. A., Kongsuwan K., Assavalapsakul W., Horwood P. F., Mitter N. and Mahony T. J. (2009) Repertoire of bovine miRNA and miRNAlike small regulatory RNAs expressed upon viral infection. PLoS One 27: 4(7):e6349.

13- 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.

14- Griffiths-Jones S. (2004) The microRNA registry. Nucleic Acids Research 32: D109–D111.

15- Griffiths-Jones S., Saini H. K., Dongen S. and Enright A. J. (2008) miRBase: tools for microRNA genomics. Nucleic Acids Research 36: D154–D158.

16- Griffiths-Jones S., Grocock R. J., Dongen S., Bateman A. and Enright A. J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Research 34: D140–D144.

17- Gu Z., Eleswarapu S. and Jiang H. (2007) Identification and characterization of microRNAs from the bovine adipose tissue and mammary gland. FEBS *Letters* 581: 981–988.

18- Hall T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. NUCLEIC ACIDS SYMPOSIUM SERIES 41: 95–98.

19- Hossain M. M., Ghanem N., Hoelker M., Rings F., Phatsara C., Tholen E., Schellander K. and Tesfaye D. (2009) Identification and characterization of miRNAs expressed in the bovine ovary. BMC Genomics 10: 443.

20- Huang T. H., Zhu M. J., Li X. Y. an Zhao S. H.

(2008) Discovery of porcine microRNAs and profiling from skeletal muscle tissues during development. PLoS ONE 3(9): e3225.

21- Huang Y., Zou Q., Tang S. M., Wang L. G. and Shen X. J. (2010) Computational identification and characteristics of novel microRNAs from the silkworm (Bombyx mori L.). Molecular Biology Rapports 37: 3171–3176.

22- Jin W., Li N., Zhang B., Wu F., Li W., Guo A. and Deng Z. (2008) Identification and verification of microRNA in wheat (*Triticum aestivum*). Journal of Plant Research 121: 351–355.

23- Kidner C. A. and Martienssen R. A. (2005) The developmental role of microRNA in plants. Current Opinion in Plant Biology 8: 38 - 44.

24- Lagos-Quintana M., Rauhut R., Lendeckel W. and Tuschl T. (2001) Identification of novel genes coding for small expressed RNAs. Science 294: 853–858.

25- Lagos-Quintana M., Rauhut R., Yalcin A., Meyer J., Lendeckel W. and Tuschl T. (2002) Identification of tissue-specific microRNAs from mouse. Current Biology 12: 735–739.

26- Lai E. C., Tomancak P., Williams R. W. and Rubin G. M. (2003) Computational identification of Drosophila microRNA genes. Genome Biology 4: R42.

27- Larkin M. A., Blackshields G., Brown N. P., Chenna R., McGettigan P. A., McWilliam H., Valentin F., Wallace I. M., Wilm A., Lopez R., Thompson J. D., Gibson TJ, Higgins DG. (2007) ClustalW and ClustalX version 2. Bioinformatics 23: 2947–2948.

28- Lee R. C., Feinbaum R. L. and Ambros V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75: 843–854.

29- Lewis B. P., Burge C. B. and Bartel D. P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120: 15–20.

30- Long J. E. and Chen H. X. (2009) Identification and Characteristics of Cattle MicroRNAs by Homology Searching and Small RNA Cloning. Biochemal Genetics 47: 329–343.

31- Pillai R. S., Bhattacharyya S. N. and Filipowicz W. (2007) Repression of protein synthesis by miRNAs: how many mechanisms? Trends in cell biology 17: 118–126.

32- Place R. F., Li L. C., Pookot D., Noonan E. J. and Dahiya R. (2008) MicroRNA-373 induces expression of genes with complementary promoter sequences. Proceedings of the National Academy of Sciences of the United States of America 105: 1608–1613.

33- Plasterk R. (2006) Micro RNAs in animal development. Cell 124: 877–881.

34- Ruby J. G., Jan C., Player C., Axtell M. J., Lee W., Nusbaum C., Ge H. and Bartel D. P. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell 127: 1193–1207.

35- Sheng X., Song X., Yu Y., Niu L., Li S., Li H., Wei C., Liu T., Zhang L. and Du L. (2010) Characterization of microRNAs from sheep (Ovis aries) using computational and experimental analyses. Molecular Biology Rapports 38: 3161–3171.

36- Singh J. and Nagaraju J. (2008) In silico prediction and characterization of microRNAs from red flour beetle (*Tribolium castaneum*). Insect Molecular Biology 17: 427–436.

37- Smith T. P. L. (2007) Discovery and profiling of bovine microRNAs from immune-related and embryonic tissues. Physiological genomics 29: 35–43.

38- Vasudevan S., Tong Y. and Steitz J. A. (2007) Switching from repression to activation: microRNAs can up-regulate translation. Science 318: 1931–1934.

39- Zhang B., Pan X. and Anderson T. A. (2006) Identification of 188 conserved maize microRNAs and their targets. FEBS Letters 580: 3753–3762.

40- Zuker M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31: 3406–3415.