International Journal of Data Science and Analysis
2016; 2(2): 21-31
http://www.sciencepublishinggroup.com/j/ijdsa
doi: 10.11648/j.ijdsa.20160202.12

Identification and Function Analysis of Novel microRNAs by Computers in *Capra Hircus*

Zhibin Ji, Guizhi Wang, Fei Dong, Lei Hou, Zhaohua Liu, Tianle Chao, Jianmin Wang*

Shandong Provincial Key Laboratory of Animal Biotechnology and Disease Control and Prevention, College of Animal Science and Veterinary Medicine, Shandong Agricultural University, Taian, China

Email address:
wangjm@sdau.edu.cn (Jianmin Wang)
*Corresponding author

To cite this article:
Zhibin Ji, Guizhi Wang, Fei Dong, Lei Hou, Zhaohua Liu, Tianle Chao, Jianmin Wang. Identification and Function Analysis of Novel microRNAs by Computers in *Capra Hircus*. *International Journal of Data Science and Analysis*. Vol. 2, No. 2, 2016, pp. 21-31.
doi: 10.11648/j.ijdsa.20160202.12

Received: September 6, 2016; **Accepted:** December 9, 2016; **Published:** December 30, 2016

Abstract: MicroRNAs are a class of non-protein coding small RNAs that regulate genes expression at post-transcriptional levels. Increasing evidence indicates miRNAs play key roles in a broad range of biological processes. In this study, based on the phylogenetic conservation of microRNAs, a combined bioinformatics and sequences homology comparison approach was used for the identification and function analysis of novel miRNA candidates in *Capra hircus*. As a result, a total of 13 potential microRNA candidates were detected following a range of filtering criteria. 153 non-redundant presumable target genes were predicted in *Ovis aries* 3′-Untranslated region database. 149 protein sequences were mapped by BLASTX, 2,517 GO terms were returned and distributed in biological process, molecular function and cell component. 66 KEGG pathways were also involved by these novel miRNAs. The qRT-PCR based assay was performed to validate the authenticity of these novel miRNA candidates. The results indicate the expressed sequence tags analysis is an efficient and affordable approach for identifying novel microRNA candidates, and our study provides insight into the further researches of miRNAs and their functions in *Capra hircus*.

Keywords: MicroRNA, Target Gene, Capra Hircus, EST, GO, KEGG, Blast2GO

1. Introduction

MicroRNAs (miRNAs) are a newly discovered family of extensively endogenous non-coding small RNA molecules, approximately 18~25 nucleotides (nt) in length. Mature miRNAs are typically generated via a two-step processing pathway. Primary miRNAs (pri-miRNA) transcribed by polymerase II or III are processed by the nuclear enzyme Drosha to produce precursor miRNA (pre-miRNA) with a characteristic hairpin structure of around 70 nt in size, which are then exported into the cytoplasm and further processed into mature miRNAs by Dicer [1]. Finally, mature miRNAs are predominantly incorporated in the RNA-induced silencing complex (RISC) in which they negatively regulate gene expression by inhibiting gene translation or degrading coding mRNA by perfect or imperfect complement to target mRNAs [2]. Increasing amounts of evidence shows that miRNAs constitute a significant group of post-transcriptional regulators, which can regulate the expression of target genes by binding to complementary sites in 3′-untranslated regions (3′-UTRs) of target genes, and extensively involved in almost all biological and metabolic processes, such as cell proliferation and apoptosis [3], viral defense [4], environment stress [5], tumorigenesis [6], and the morphogenesis of specific organs [7]. There are several methods for identification of miRNAs, such as biochemical method, direct cloning and sequencing method and bioinformatics prediction method. The genetic screening method is initially used in identification of lin-4 and let-7 in *Caenorhabditis elegans* (C. *elegans*) [8, 9]. Soon afterwards, the direct cloning method and sequencing was also used for the discovery of novel miRNAs [10]. But these methods have several disadvantages, they are expensive, time consuming, inefficient and difficulty to find miRNAs expressed in specific tissues or at specific times. More interestingly, many studies have found that many miRNAs are highly evolutionarily conserved among different organisms, the miRNA genes in one species may have homologues or orthologues in other species, and this conservation ranges a long distance, such as from worms to humans in animals and from moss and ferns to high

flowering plants [11, 12]. Following the development of computational biology, bioinformatics and genomics, this evolutionarily conserved property of the miRNAs provides a powerful strategy to identify homologue miRNAs in a new animal species using already known miRNAs in other species through the use of a comparative genome-based homologue search [13, 14]. Although it is deficient in the genome information of *Capra hircus*, published Expressed Sequence Tags (ESTs) databases in GenBank have made it possible to obtain more genetic information. The ESTs are partially expressed gene sequences that have been converted into cDNAs, and the EST analysis is most well-known in the bioinformatics strategy [15]. Up to now, the EST analysis based on computational or bioinformatics approach to identify novel miRNAs which are expressed only at a certain developmental stage, specific tissues, or at less copy number that was not possible with other approaches of sequencing and cloning, has been proven to be an economical, feasible, and fast method for gene discovery in species lacking a draft genome sequence [16]. Based on some strict parameter setting in computer tools, using this method, many of miRNAs have been systematically identified in both plants, such as soybean [17], wheat [18], tobacco [19], potato [20], cotton [21], catharanthus roseus [22], coffee [23] and blueberry [24].

Capra hircus is one of the most important agricultural livestock for meat and milk production that is raised broadly in the world and very popular in many countries. More important, it is also an ideal model organism for biological and comparative genomics studies. Although miRNAs have been extensively studied in the past few years, no systematic study has been performed on *Capra hircus*, there is limited information compared with other mammals, and only 436 mature miRNAs were presented in miRBase v21. So much effort is still needed to find potential novel miRNA genes in the *Capra hircus* to better understand their roles in the regulation of gene expression. Thus, in this study, we are intended to predict and discover more miRNAs with EST strategy and to analyze their target genes. Based on our findings, 13 miRNAs were identified for the first time in *Capra hircus*. Furthermore, 153 putative target genes for these newly identified miRNAs were also identified. GO annotation and KEGG pathways analysis shown that the putative target genes are involved in various biological processes, such as signal transduction, metabolism, and development.

2. Materials and Methods

2.1. Sequence Datasets

The *capra hircus* ESTs (expressed sequence tags, total 14,479 sequences) were downloaded from the NCBI GenBank nucleotide database (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/nucest?term=est%20goat). The non-redundant 3´-UTRs of *Ovis aries* were downloaded from UTRdb database (http://utrdb.ba.itb.cnr.it/) [25] for target genes prediction. All available animal miRNAs and their precursor sequences, totally 18,698 mature miRNAs and 15,828 pre-miRNAs in 97 species, were downloaded from miRBase (Relase 20, August 2012 at http://www.mirbase.org/) [26]. miRNAs from *Bos taurus* and *Ovis aries* (closest homolog of *Capra hircus*, 858 mature miRNAs and 821 pre-miRNAs) were present as known miRNAs of *Capra hircus*. The homologous miRNAs were eliminated and the remaining sequences were defined as reference database and were compared with the local ESTs for searching *Capra hircus* new miRNA candidates. CD-HIT software [27] was used to trim out the redundant sequences and created local database for miRNAs and ESTs.

Computational identification of Capra hircus miRNA

Comparative software BLAST-2.2.27 was downloaded from NCBI, the online BLASTX (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&BLAST_PROGRAMS=blastx&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome) was used to search for protein coding sequences. EST contigs downloaded from NCBI GenBank database were used as a local database to BLAST against the local miRNAs dataset to identify potential novel miRNA candidates with a expect value =1,000. Mfold (http://mfold.rna.albany.edu/?q=mfold) [28] was used to analyze the secondary structure of RNAs. A flow chart of our prediction procedure for searching conserved *Capra hircus* miRNA homologues is summarized in figure 1. Secondary structures of resulting ESTs were calculated and their MFE scores were recorded. A six-step prediction method was used to identify *Capra hircus* miRNAs. First, alignment itself of known miRNAs of animal was conducted by CD-HIT Suite to remove redundant sequences with similarity ≥80%. Second, we used remaining miRNAs as a query sequences database for local BLAST searched against *Bos Taurus* and *Ovis aries* miRNAs, which had the highest conserved genomic information with *Capra hircus*, to remove the homologue miRNAs. Third, the remaining miRNAs as second query sequences database for local BLAST searched against the *Capra hircus* ESTs. The parameter settings were as follows: E-value cut-off was 1, the number of descriptions and alignments were 1000. Fourth, the obtained ESTs which have >90% similarity (pairing number ≥50, mismatch number ≤2) with the corresponding known pre-miRNAs (precursor miRNAs) sequences were selected to the BLAST searched against the protein database at NCBI by BLASTX online program to remove protein-coding sequences. The parameters were set to default value. The last step was to apply Mfold program [28] to further identify the pre-miRNAs. Four criteria were used according to Zhang *et al*. [29]: a) the A+U content of the precursor sequences should range from 30~70%; b) the mature miRNAs should locate on one arm of the hairpin structure; c) the minimum free energy of the secondary structure for each potential pre-miRNA was less that -18 Kcal/mol; d) the hairpin must include at least 16 bp within the first 22 nt of the miRNA; e) miRNA had less than seven mismatches with the opposite miRNA sequence in the other arm; f) not loops or bulges in miRNA sequences, particularly not large asymmetric bulges.

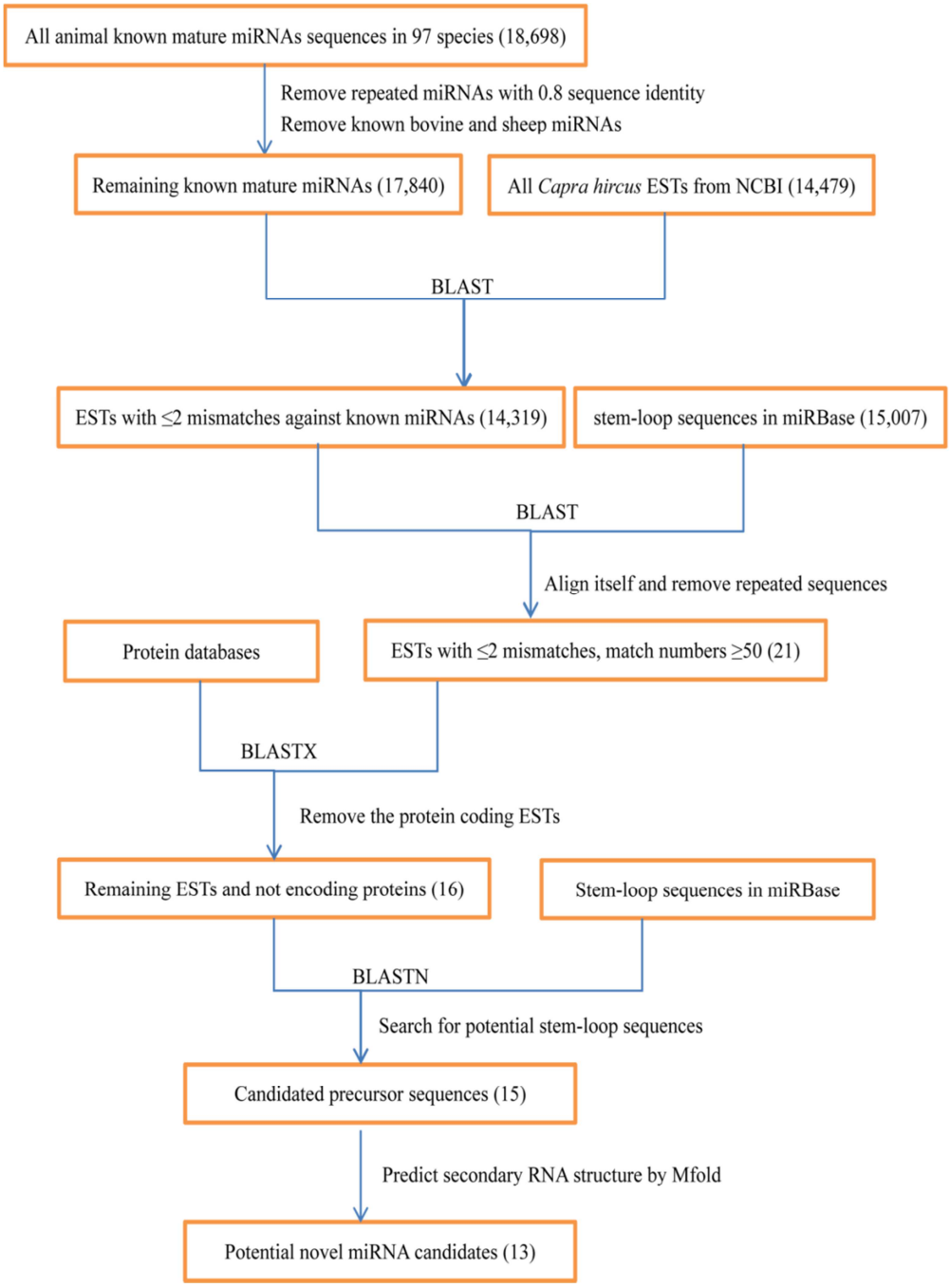

Figure 1. *The flow chart of potential novel miRNA candidates search. The prediction was performed by homologous BLAST between the ESTs database of Capra hircus and previously known animal miRNAs in miRBase 20.*

2.2. Target Genes Prediction for Novel miRNA Candidates

As for *Capra hircus*, because only few gene sequences available and no 3′-UTR database is current available, we used 3′-UTR database of *Ovis aries* as a reference system for finding the target genes of the novel miRNA candidates. The targets were then predicted with the help of RNAhybrid 2.2 [30], the parameter settings are as follows: a) Helix constraint (Seed match region) is 2~7; b) MFE percentage is 75%; c) Energy cut-off is -25Kcal/mol; d) one G:U pair in the seed is allowable; e) the size of max internal loop and bulge loop is 2. This program is based on the criteria suggested by Allen et al. [31] and Schwab et al. [32]. Briefly, the criteria are as follows: a) no more than four mismatches between the sRNA and the target (G-U bases count as 0.5 mismatches); b) no more than two adjacent mismatches in the miRNA/target duplex; c) no adjacent mismatches in positions 2-12 of the miRNA/target duplex (5′ of miRNAs); d) no mismatches in positions 10-11 of the miRNA/target duplex; e) no more than 2.5 mismatches in positions 1-12 of the miRNA/target duplex (5′ of miRNAs); and f) MFE of the miRNA/target duplex should be ≥75% of the MFE of the miRNA bound to its perfect complement.

2.3. GO Annotations and KEGG Pathways Analysis for Target Genes

Gene ontology analyses were carried out with Blast2GO (http://www.Blast2GO.com/b2ghome) [33] at AmiGO (http://amigo.geneontology.org/cgi -bin/amigo /go.cgi) [34] from three categories: molecular function, cellular component and biological process. The 500 MB option of Blast2GO was installed, the calculations consist of three key sequential steps: a) BLAST, b) Mapping, and c) Annotation. In BLAST, FASTA-formatted sequences of target genes were submitted to the BLASTX server at the NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&BLAST_PROGRAMS=blastx&PAGE_TYPE=Bla-stSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome) to query the protein database, and generated hits and gene names/accessions. The parameter settings were as follows: Number of Blast hits is 20, Blast Expect Value is 0.001. The mapping algorithm uses the parameters of the Blast Table to search various databases to identify and retrieve Gene Ontologies (GO) associated with the hits obtained from NCBI BLAST searches. The annotation procedure selects the GO terms from the GO pool obtained by the Mapping step and assigning them to the query sequences. Every GO annotation must provide valid evidence, known as Evidence Codes, which encompass a broad range of empirical or other support such as electronic annotation or direct assay. The results of GO analysis are presented in the form of Directed Acyclic Graph (DAG), a DAG is a hierarchical representation of Ontology terms in a way that depicts the directional relationships between parent-child GO term nodes. Enzyme codes were also mapped with Blast2GO and downloaded from KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/) [35] to understand their biological processed involved by novel miRNAs.

2.4. Quantitative Real-Time RT-PCR

miRNA was isolated using the miRcute miRNA isolation kit (TIANGEN, DP501) according to the manufacturer's protocol. We examined the quality of RNA using an Agilent 2100 Bioanalyzer and stored at -80°C for further use. Real-time quantitative PCR was performed with SYBR® Premix Ex Taq™ II (TaKaRa Biotechnology Co., Ltd., Japan, DRR081A) by Mx3000p™ SYBR® Green real-time quantitative PCR Analyzer (Stratagene, USA). Briefly, 2 μg of miRNA was reverse transcribed using the One Step PrimeScript® miRNA cDNA Synthesis Kit (TaKaRa Biotechnology Co., Ltd., Japan, D350A). The reverse transcription reaction system included 10 μL of 2X miRNA Reaction Buffer, 2 μL of 0.1% BSA, 2 μL of miRNA PrimeScript® RT Enzyme Mix, 2 μL of total RNA (10 pg/μL~1 μg/μL), and RNase-Free dH_2O to a final volume of 20 μL. The RT-PCR program was set to 37°C for 60 min followed by 85°C for 5 sec. The cDNA products were stored at -20°C. The reaction solution was prepared on ice, and comprised 10 μL of 2X SYBR® Premix Ex Taq™ II, 0.8 μL of PCR Forward Primer (10 μM), 0.8 μL of Uni-miR qPCR Primer (10 μM), 0.4 μL of 50X ROX Reference Dye II, 2 μL of cDNA, and dH_2O to a final volume of 20 μL. The reaction mixtures were incubated in a 96-well plate at 95°C for 30 sec, followed by 40 cycles of 95°C for 5 sec, 60°C for 30 sec and 72°C for 30 sec. All reactions were performed in triplicate. The primers for the miRNAs had the same sequences as the novel miRNA candidates with appropriate adjustments at their 5' terminus (Table 1).

Table 1. *Summary of miRNA primers used in real-time RT-PCR.*

Primer	Sequences (5′→3′)	Length (nt)	GC (%)	Tm
U6	CAAGGATGACACGCAAATTCG	21	47.6	69
miR-4426	GGCGGAAGATGGACGTACTTT	21	52.4	63
miR-4680-5p	CGAGAACTCTTGCAGTCTTAGATGT	25	44	61
miR-4680-3p	GCGGTCTGAATTGTAAGAGTTGTTA	25	40	61
miR-5047	GCAGCTGCGGTTGTAAGGT	19	57.9	61
miR-3064-5p	TCTGGCTGTTGTGGTGTGC	19	57.9	62
miR-3064-3p	GCCACACTGCAACACCTTACA	21	52.4	62
miR-3661	CTGGGACTCGGACAGCTG	18	66.7	61
miR-6527	CGTGGACGAAGAGATGGGA	19	57.9	62
miR-1244	GCGTAGTTGGTTTGTATGAGATGGTT	26	42.3	64
miR-2904	ATATAAGCCTCGGTCGGCCTC	21	57.1	64
miR-2887	AATATACGGGGTCCGGTGCG	20	60	66
miR-716a	ACGGTGAGCCTTGAAGCCT	19	57.9	62
miR-716b	CGCGTGGTGGTAGTAGCAAATAT	23	47.8	62

3. Results

3.1. Identification and Characteristics of Potential miRNA Candidates

To exploit the new miRNA candidates of *Capra hircus*, a computational approach is used to search homologous sequences in *Capra hircus* ESTs database. After removing the identical or highly similar sequences (80% similarity) in animal miRNAs database, getting rid of the known miRNAs of *Bos taurus* and *Ovis aries*, 17,840 mature miRNAs and 15,007 pre-miRNAs were remained, as local miRNAs database respectively, for further BLAST to local ESTs database. Following the strict filtering criteria (figure 1), we finally identified 13 potential miRNA candidates of *Capra hircus*, and the detailed characteristics such as sequences, location on genome, size and minimal folding free energies of precursor sequences and A+U content are tabulated in Table 2. Among which, 7 mature miRNAs were located at the 5′ arm of

the precursor, and the other 6 were located at the 3′ arm. 4 were located in the sense strand and the rest were in antisense strand. 10 newly predicted miRNAs were perfectly matched with the corresponding homologues, whereas the remaining 3 mature miRNA sequences (miR-3661, miR-2904, miR-716a) differ by 1~2 nucleotides from their homologues.

Table 2. *The characteristics of 13 predicted novel miRNA candidates in Capra hircus.*

miRNA	Sequence	GenBank (ID)	Size (nt)	NM (nt)	SP (nt)	δ-len (nt)
miR-4426	GAAGAUGGACGUACUUU	1292728 31088741	17	0	63	0
miR-4680-5p	AGAACUCUUGCAGUCUUAGAUGU	152114074	23	0	66	3
miR-4680-3p	UCUGAAUUGUAAGAGUUGUUA	152114074	21	0	66	3
miR-5047	UUGCAGCUGCGGUUGUAAGGU	152114634 152114644	21	0	101	1
miR-3064-5p	UCUGGCUGUUGUGGUGUGCAAA	152114634 152114644 152116412	22	0	67	0
mir-3064-3p	UGCCACACUGCAACACCUUACA	152114634 152114644 152116412	22	0	67	0
miR-3661	UCACCUGGGACUCGGACAGCUG	152116126 152125556	22	1	96	0
miR-6527	CACGUGGACGAAGAGAUGGGA	152117335	21	0	95	0
miR-1244	AAGUAGUUGGUUUGUAUGAGAUGGUU	152121951 152121952	26	0	86	1
miR-2904	GGGAGCCUCGGUCGGCCUC	152123978	19	1	70	0
miR-2887	CGGGACCGGGGUCCGGUGCG	152123978	20	0	57	0
miR-716b	AGAUCUUGGUGGUAGUAGCAAAUAU	152124040	25	0	114	0
miR-716a	GCGGUGAGCCUUGAAGCCU	152124040	19	2	149	0

Table 2. *Continued.*

miRNA	Sequence	GenBank (ID)	A+U (%)	Strand	NT	MFE (kcal/mol)
miR-4426	GAAGAUGGACGUACUUU	1292728 31088741	61.9	-	0	-14.5
miR-4680-5p	AGAACUCUUGCAGUCUUAGAUGU	152114074	70.77	-	1	-21.4
miR-4680-3p	UCUGAAUUGUAAGAGUUGUUA	152114074	70.77	-	0	-21.4
miR-5047	UUGCAGCUGCGGUUGUAAGGU	152114634 152114644	52.48	+	21	-30.3
miR-3064-5p	UCUGGCUGUUGUGGUGUGCAAA	152114634 152114644 152116412	52.24	+	62	-23.8
mir-3064-3p	UGCCACACUGCAACACCUUACA	152114634 152114644 152116412	52.24	+	4	-23.8
miR-3661	UCACCUGGGACUCGGACAGCUG	152116126 152125556	42.71	-	31	-32.8
miR-6527	CACGUGGACGAAGAGAUGGGA	152117335	69.23	+	4	-22.3
miR-1244	AAGUAGUUGGUUUGUAUGAGAUGGUU	152121951 152121952	68.6	-	13	-15.8
miR-2904	GGGAGCCUCGGUCGGCCUC	152123978	17.14	-	77	-39.6
miR-2887	CGGGACCGGGGUCCGGUGCG	152123978	15.79	-	74	-36.7
miR-716b	AGAUCUUGGUGGUAGUAGCAAAUAU	152124040	38.93	-	16	-41.7
miR-716a	GCGGUGAGCCUUGAAGCCU	152124040	38.93	-	13	-51.1

NM: the mismatch number of identified miRNA sequences compared to those of known miRNAs in related species. SP: Sequence length of pre-miRNAs. δ-len: Dislocation number of mature miRNAs in their pre-miRNAs compared to their homologues/orthologues in related species. +: Sense strand. -: Antisense strand. NT: Number of predicted target genes. MFE: Minimal fold free energy of pre-miRNAs.

The mature miRNA sequences predicted were 17~26 nt in length centering on 20~22 nt. This is coincident with the specificity of Dicer processing [36]. The length of these newly identified miRNA precursors varied from 57~149 nt with an average of 84 nt. The A+U content of 9 miRNAs were evaluated ranging from 38.93% to 70.77%, which is in agreement with the previous results [37], but 2 miRNA (miR-2904, miR-2887) have lower A+U content than 30% (Table 2). The miRNA precursors, unlike other non-coding RNAs, have lower folding free energy than random sequence. Minimal folding free energy is one of the important features to identify new miRNA genes [38]. The analysis of these newly identified miRNAs precursors shown that the minimal folding free energies ranges from -14.5~-51.1 Kcal/mol with an average of -28.9 Kcal/mol (Table 2). Two miRNAs (miR-4426, miR-1244) have higher MFE than -18 Kcal/mol, these is similar as their homologs in other species.

3.2. Target Genes Prediction for Novel miRNA Candidates

To further understand the physiological functions and biology processes involved by these miRNAs in *Capra hircus*, target gene prediction was performed using RNAhybriad 2.2 [30]. Based on the criteria described in "Materials and methods", using the newly identified miRNA sequences, we scanned the 3′-UTR database of *Ovis arise*, a total of 153 non-redundant potential target genes were identified for 11 novel miRNA candidates based on their complementarity with their target sequences (Table 2 and S 1). As seen in Table 2, each miRNA has multiple target genes, there are 77 target genes for miR-1904, and only one target gene for miR-4680-5p, no target gene was detected for two miRNAs (miR-4426, miR-4680-3p) in our study, may be due to the imperfect 3′-UTRs database of *Ovis aries*.

3.3. GO Annotation and KEGG Pathway Analysis of Target Genes

To better understand the biological functions and molecular interaction networks involved by these novel miRNAs, GO annotation and KEGG pathway analysis were performed in our study. Firstly, we blasted 153 target gene sequences to protein databases at NCBI with BLASTX, 151 sequences were retrieved and used for GO annotation and KEGG pathway analysis (figure 2: A), the corresponding hit distribution in species was shown in figure 2: B. Secondly, 151 sequences were mapped on AmiGO database, and 149 sequences were retrieved with GO terms, a total of 2,517 ontology terms were returned (figure 2A and S 2). Of which, 140 target gene sequences were assigned to the GO terms of "biological process" ontology, of which, "cellular process", "metabolic process" and "biological regulation" were over-represented, occupied for 69.29%, 67.14% and 66.43%, respectively. 144 sequences were assigned to the "molecular function" ontology, most of which were found to take part in the functions of "binding" (84.03%) and "catalysis" (41.67%). 143 sequences were assigned to the "cellular component", and mainly distributed in cell (90.91%), organelle (57.69%), and/or membrane (49.23%) (figure 3 and S 3). KEGG pathway analysis showed that 118 target genes were distributed to 66 biological pathways. Most of these genes were involved in cellular metabolism, biosynthesis, diseases and signal transduction (S 4). The most commonly pathway was the metabolic pathway, with 57 genes representing 48.31% of the total. Further analysis for target genes is needed to help us gain insight into the roles of these newly identified miRNAs in *Capra hircus*.

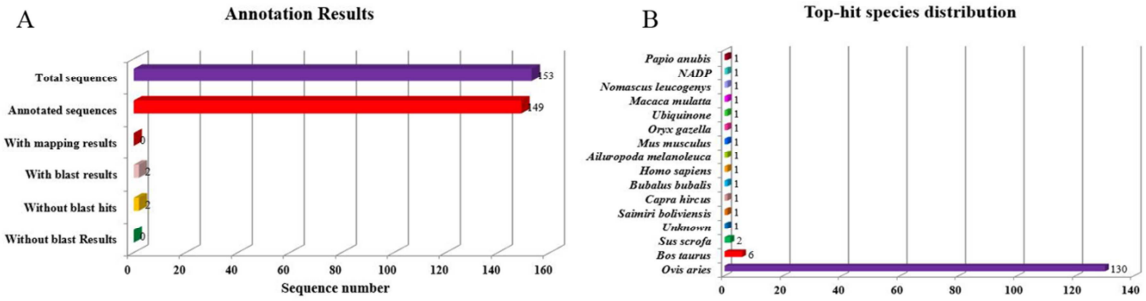

Figure 2. *The results of mapping and annotation at AmiGO database. A: Results of annotation at AmiGO database. For 153 targets genes, 149 were annotated, 2 were not blasted at nucleotide database. B: Distribution of Top-hit sequences in species based on BLASTX. Most sequences distributed in Ovis aries species, followed by Bos taurus, only 15 sequences distributed in other species.*

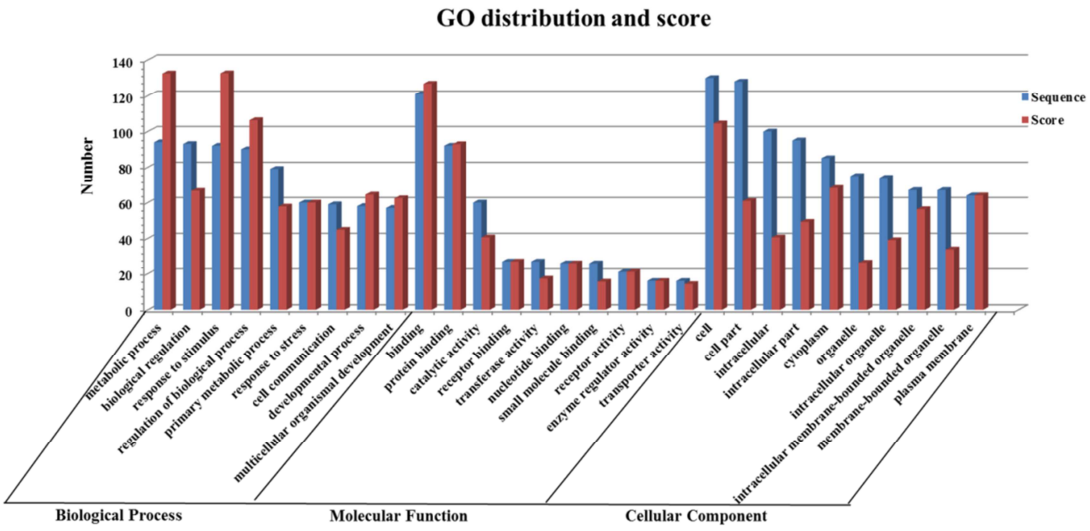

Figure 3. *The distribution of part GO terms at three categories: biological process, molecular function, cellular component. A total of 2,517 ontology terms were assigned for 149 target genes, 140 target gene sequences were annotated to "biological process" ontology, 144 sequences were annotated to the "molecular function" ontology, 143 sequences were annotated to the "cellular component".*

3.4. Validation of Potential miRNAs by qRT-PCR

To validate the novel miRNA candidates predicted in our study, we performed a qRT-PCR based assay to support the existence of these miRNAs. Mammary gland tissues were used for total RNA extraction. Each sample was replicated for three times. By this approach, 11 predicted miRNAs were confirmed (figure 4), and 2 miRNAs (miR-6527, miR-716a) were not detected under the conditions of the present study, which require further validation. The reason for it may be result from the facts: a) the abundance of the missed miRNAs might be too low to be detected in the tissues; b) they might be inducible miRNAs whose expression is under the control of some factors; and c) expression of these miRNA may be time- or tissue-specific. d) the primers for them are not suitable.

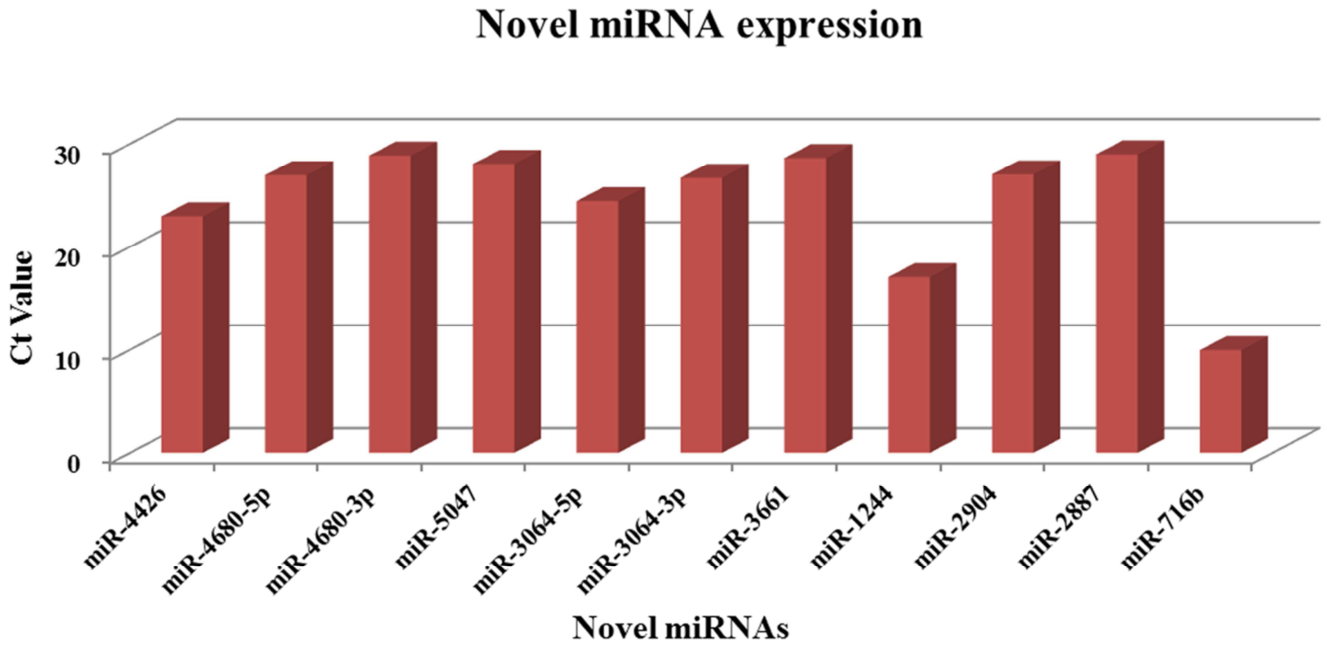

Figure 4. *qRT-PCR validation of the identified novel miRNA candidates. Total RNA was extracted respectively from mammary gland tissues and used for qRT-PCR, the relative expression abundance were expressed as Ct value, each sample was replicated for three times.*

4. Discussion

miRNAs are single-stranded non-protein coding small RNA regulating target mRNAs expression at the post-transcriptional levels [39]. miRNAs are evolutionarily conserved from species to species within the same kingdom, and miRNAs in one species may exist as orthologs or homologues in other species [40]. This suggests a powerful strategy for bioinformatics identification of the homologues or orthologs of miRNA genes in other species. Currently, a large number of miRNAs were identified through a series of method and deposited in miRBase. With the availability of sequence resources in public databases, miRNA identification methods based computational homology search are increasingly concerned in recent years due to its advantages of feasible operation, low cost and high efficiency. At present, several databases, such as Genome, GEO (Gene Expression Omnibus), GSS (Genome Survey Sequences) and EST are mainly used for miRNA mining. Considering the unavailability of genome and genomic survey sequences of *Capra hircus*, ESTs database was mined for the identification of miRNAs [41]. Sequence and structure homology are the main theory behind the computer-based approach for miRNAs prediction [13, 40, 42]. In this study, using computer-based homologous sequences search, we found 13 novel miRNA candidates, and the results were verified by qRT-PCR (figure 4), and these miRNAs (miR-4426, miR-3064-5p, miR-3064-3P, miR-6527, miR-1244, miR-716b, miR-1887, miR-2904) were also found in our previous studies by high-throuthput sequencing [43, 44]. The miRNA precursors exhibited various sizes in length and MFE, 13 identified miRNA precursors could be folded into the typical secondary structure of miRNAs one example are shown in figure 5 (all miRNA secondary structure are also shown in S 5). MFE is one of the important features to identify new miRNA genes [45], unlike other noncoding RNAs, miRNAs have lower MFE than random sequence. The analysis by Mfold revealed these newly identified miRNA precursors have negative minimal folding free energies lower than -18 Kcal/mol (except for miR-4426 and miR-1244, they are similar to their homologs in other species), implying that these miRNA precursors are positive.

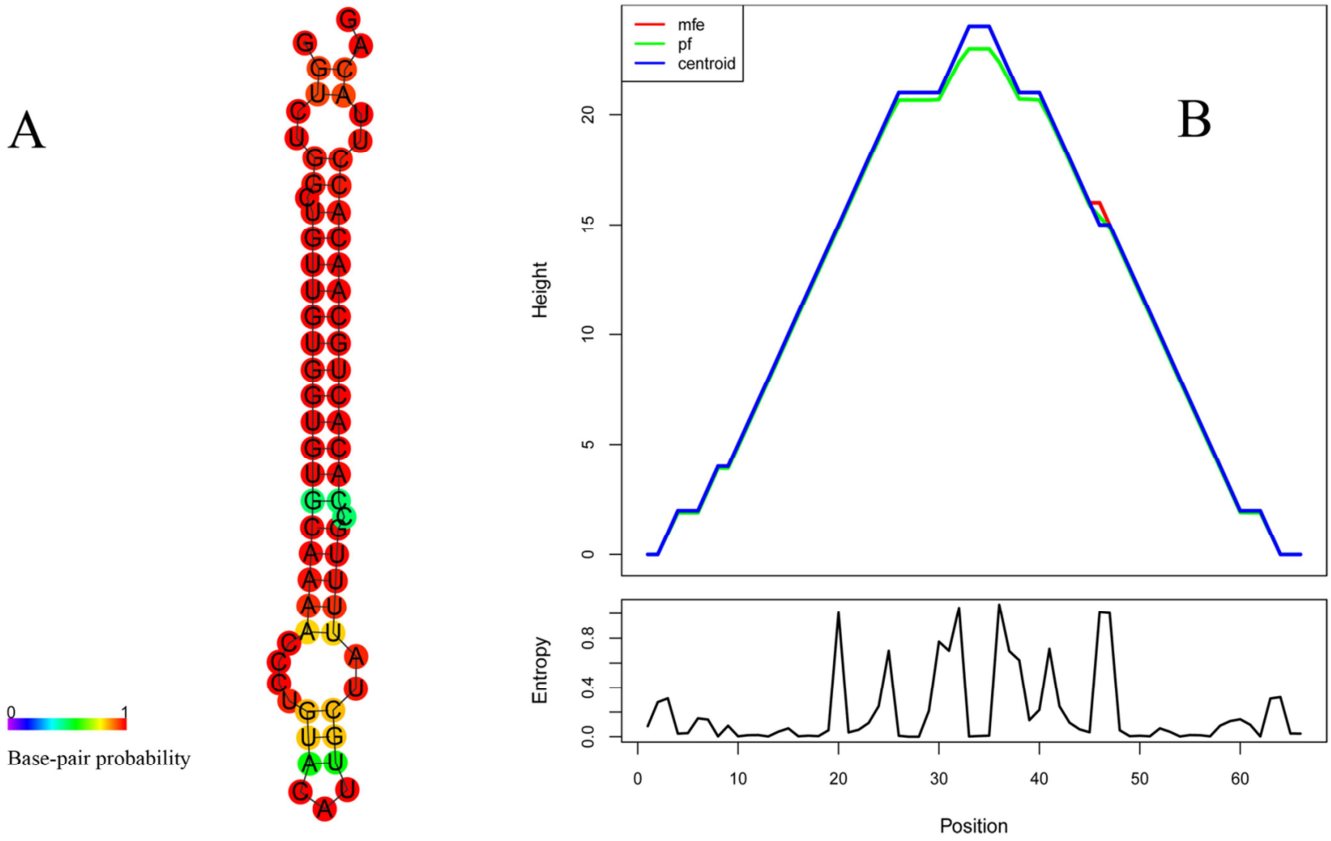

Figure 5. *The secondary structure characteristics predicted for miR-3064-3p using RNAfold software. A: the interactive drawing of the MFE structure based on the base-pair probabilities. B: a mountain plot representation of the MFE structure, the thermodynamic ensemble of RNA structures, and the centroid structure, the positional entropy for each position was also presented.*

The miRNA target gene identification is an important step for understanding the role of miRNAs in gene regulatory networks. miRNAs can regulate gene expression by binding to complementary sites on the target mRNAs, usually located in the 3′-UTR. It was found that conserved motifs of 3′-UTR were likely under regulation mediated by miRNAs [46]. In our study, we used the conservative 3′-UTRs of *Ovis aries* for target genes prediction, which make our results more convictive than previous studies, which used mRNA sequences for target genes prediction. Based on the complementarity between the seed region and 3′-UTR and the strict parameter settings, a total of 153 potential target genes were identified for the 11 identified miRNA candidates (two miRNAs, miR-4426 and miR-4680-3p, have no results, S 1). By analysis of these target genes, we found that most target genes encode transcription factors that regulate individual development, signaling transduction, metabolism, and binding functions. miR-4680-5P targeted SLC25A19, which is an important regulatory protein controlling development [47]. miR-1244 targeted IL-10, which mainly encoded a cytokine possessing pleiotropic effects in immune-regulation and inflammation, and is involved in the regulation of the JAK-STAT signaling pathway [46]. miR-2094 targeted PANK2, which is a key regulatory enzyme in the biosynthesis of coenzyme A (CoA) in mammalian cells. It catalyzes the first committed step in the universal biosynthetic pathway leading to CoA and is itself subject to regulation through feedback inhibition by acyl CoA species [48].

The analysis of target genes revealed that more than one gene was regulated by individual miRNAs, which suggested that miRNA research should be focused on networks rather than individual connections between miRNA and strongly predicted targets [49, 50]. GO (Gene Ontology) annotation and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis are essential tool to understand the functions of target genes. GO is a universal standard terminology used for unifying the representation of gene and gene product attributes across all species, the annotations are classified from three categories: biological process, molecular function and cellular component. For GO annotations, as discussed in the Materials and methods, with critical parameter settings: E-value hit filter is $1.0E^{-6}$, Annotation cut-off is 55, GO weight is 5, we observed various degrees of GO terms from various databases at AmiGO (figure 6: A), and the corresponding Evidence code were shown in figure 6: B. In our study, we also observed GO enrichment, such as in the biological process, most terms were enriched at 6, 7 and 8 GO levels. In molecular function and cellular component, 4, 5 and 6 GO levels were more enriched (figure 7). The results show that unambiguous annotations were accurately assigned to each target genes. KEGG is a database resource for understanding high-level functions and utilities of the biological system from molecular-level information. KEGG pathways are used to illustrate the molecular interaction networks of genes and/or

proteins. KEGG analysis showed that approximately 48.31% of the genes were committed to a metabolic pathway, 19.50% of genes were committed to biosynthesis pathway. These results indicated that various miRNAs might be involved in physiology processes in goat. Further experiment validations are needed to verify the biological pathways involved by the miRNAs.

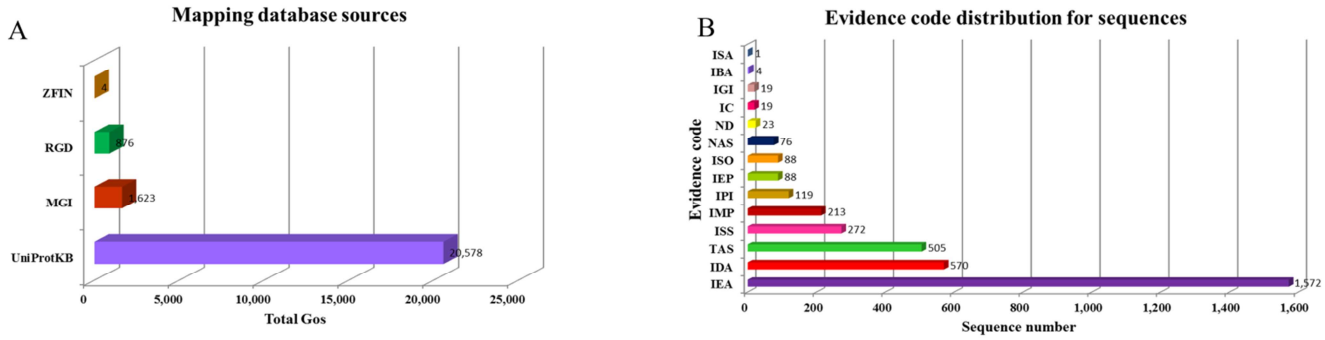

Figure 6. *The database sources and evidence code distribution of mapping on AmiGO. A: GO terms at different databases. B: Distribution of Evidence code for each GO terms by mapping. IEA: inferred from electronic annotation. IDA: inferred from direct assay. TAS: traceable author statement. ISS: inferred from sequence or structural similarity. IMP: inferred from mutant phenotype. IPI: inferred from physical interaction. IEP: inferred from expression pattern. ISO: inferred from sequence orthology. NAS: non-traceable author statement. ND: no biological data available. IC: inferred by curator. IGI: inferred from genetic interaction. IBA: inferred from biological aspect of ancestor. ISA: inferred from sequence alignment.*

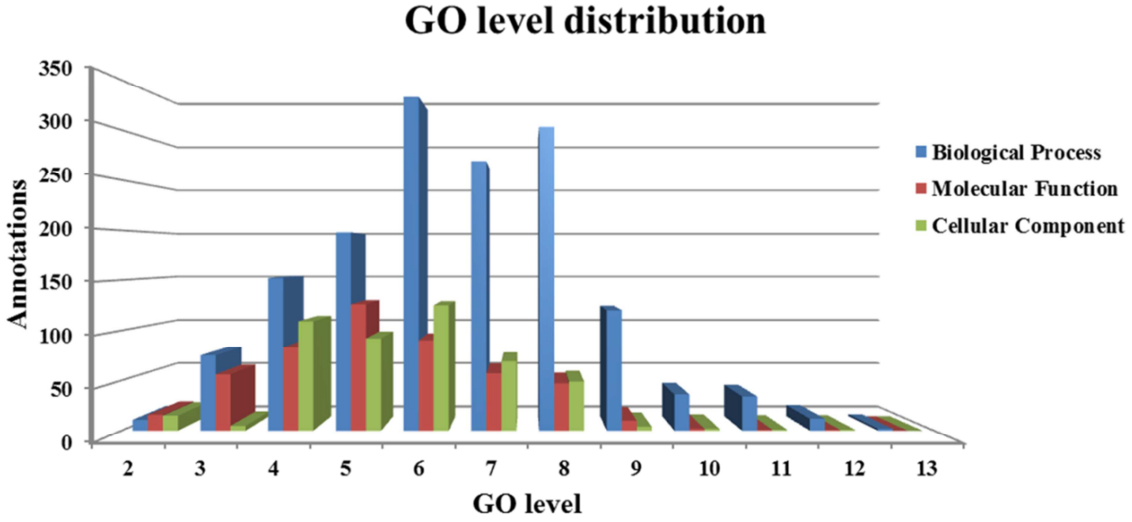

Figure 7. *GO level distribution. Number of annotations per level for all three GO categories: Biological process, Molecular function, Cellular component.*

5. Conclusion

In this study, based on local BLAST with ESTs database of *Capra hircus*, we used previously known animal miRNAs to search for new miRNAs. A total of 13 novel miRNAs were detected and validated, 153 target genes were also identified using 3′-UTRdb of *Ovis aries*. GO annotations enrichment and KEGG pathways analysis indicate that miRNAs are extensively involved in the regulation of various biological processes. The current results confirm that the approach of EST analysis is a relatively efficient means of identifying miRNAs to those species whose genomes are not available, and will pave the way for understanding the function and processing of miRNAs in future. Moreover, these findings are the good functional genomic resources for understanding the gene regulatory mechanism in *Capra hircus*.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 31672401), Shandong Provincial Natural Science Foundation of China (Grant No. ZR2014CM029), a project of Shandong Province Higher Educational Science and Technology Program of China (Grant No. J14LF07), and Shandong Provincial Modern Agriculture Industry Technology System (Grant No. SDAIT-10-16).

Appendix

S 1: Target genes predicted for 11 novel miRNA candidates. Target gene prediction was performed using RNAhybrid 2.2, the parameter settings as follows: Helix constraint (Seed match region) is 2~7. MFE percentage is 75%. Energy cut-off

is -25Kcal/mol. one G:U pair in the seed is allowable. The size of max internal loop and bulge loop is 2.

S 2: The results of GO annotation for 149 target genes at AmiGO database. A total of 2,517 ontology terms were assigned to 149 target genes.

S 3: The distribution of part GO terms at three categories: biological process, molecular function, cellular component. 140 target gene sequences were annotated to "biological process" ontology, 144 sequences were annotated to the "molecular function" ontology, 143 sequences were annotated to the "cellular component".

S 4: Pathway involved in by target genes. 118 target genes were distributed to 66 biological pathway, most of these genes were involved in cellular metabolism, biosynthesis, diseases and signal transduction.

S 5: Secondary structure predicted by Mfold software for novel miRNAs.

References

[1] BARTEL, D. P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 16: 281-297.

[2] CARRINGTON, J. C. AND AMBROS, V., 2003. Role of microRNAs in plant and animal development. *Science*, 301: 336-338.

[3] GU, Y., WANG, X. D., LU, J. J., LEI, Y. Y., ZOU, J. Y. AND LUO, H. H., 2015. Effect of mir-16 on proliferation and apoptosis in human A549 lung adenocarcinoma cells. *Int. J. Clin. Med.*, 8: 3227-3233.

[4] PEDERSEN, I. M., CHENG, G., WIELAND, S., CROCE, C. M. AND CHISARI, F. V., 2007. Interferon modulation of cellular microRNAs as an antiviral mechanism. *Nature*, 449: 919-922.

[5] LEUNY, A. K. AND SHARP, P. A., 2010. microRNA functions in stress responses. *Mol. Cell*, 40: 205-215.

[6] SINI, R. A., TRINK, B. AND NISSAN, A., 2009. The role of microRNA in tumorigenesis: key players or innocent bystanders. *J. Surg. Oncol.*, 99: 135-136.

[7] MAS, V. R., DUMUR, C. I., SCIAN, M. J., GEHRAU, R. C., MALUF, D. G., 2013. MicroRNAs as biomarkers in solid organ transplantation. *Am. J. Transplant*, 13: 11-19.

[8] 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.

[9] REINHART, B. J., SLACK, F. J., BASSON, M., PASQUINELLI, A. E., BETTINGER, J. C., ROUGVIE, A. E., HORVITZ, H. R. AND RUVKUN, G., 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403: 901-906.

[10] LAGOS-QUINTANA, M., RAUHUT, R., LENDECKEL, W. AND TUSCHL, T., 2001. Identification of novel genes coding for small expressed RNAs. *Science*, 294: 853-858.

[11] WILLMANN, M. R. AND POETHIQ, R. S., 2007. Conservation and evolution of miRNA regulatory programs in development. *Curr. Opin. Plant Biol.*, 10: 503-511.

[12] CHAVEZ, M. R. A., DE FATIMA, R. C. F., DE PAOLI, E., ACCERBI, M., RYMARGUIS, L. A., MAHALINGAM, G., MARSCH, M. N., MEYERS, B., C., GREEN, P. J., DE FOLTER, S., 2014. Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat. Commun.*, 5: 3722-3732.

[13] KIM, V. N. AND NAM, J. W., 2006. Genomics of microRNA. *Trends genet.*, 22: 165-173.

[14] LI, Y., ZHANG, Z., 2015. Computational Biology in microRNA. *Wiley Interdiscip Rev. RNA*, 6: 435-452.

[15] ADAMS, M. D., KELLEY, J. M., GOCAYNE, J. D., DUBNICK, M., POLYMEROPOULOS, M. H., XIAO, H., MERRIL, C. R., WU, A., OLDE, B. AND MORENO, R. F., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252: 1651-1656.

[16] MISHRA, A. K. AND LOBIYAL, D. K., 2011. miRNA prediction using computational approach. *Adv. Exp. Med. Biol.*, 696: 75-82.

[17] CHEN, R., HU, Z. AND ZHANG, H., 2009. Identification of MicroRNAs in Wild Soybean (Glycine soja). *Journal of Integrative Plant Biology*, 51: 1071-1079.

[18] HAN, Y. S., LUAN, F. L., ZHU, H. L., SHAO, Y., CHEN, A. J., LU, C., LUO, Y. AND ZHU, B., 2009. Computational identification of microRNAs and their targets in wheat (Triticum aestivum L.). *Science in China Series C-Life Sciences,* 52: 1091-1100.

[19] FRAZIER, T. P., XIE, F. L., FREISTAEDTER, A., BURKLEW, C. E. AND ZHANG, B. H., 2010. Identification and characterization of microRNAs and their target genes in tobacco (Nicotiana tabacum). *Planta*, 232: 1289-1308.

[20] XIE, F. L., FRAZIER, T. P. AND ZHANG, B. H., 2011. Identification, characterization and expression analysis of MicroRNAs and their targets in the potato (Solanum tuberosum). *Gene*, 473: 8-22.

[21] WANG, M., WANG, Q. AND WANG, B., 2012. Identification and characterization of microRNAs in Asiatic cotton (Gossypium arboretum L.). *PLoS One*, 7: e33696.

[22] PANI, A., MAHAPATRA, R. K., 2013. Computational identification of microRNAs and their targets in *Catharanthus roseus* expressed sequence tags. *Genomics Data*, 1: 2-6.

[23] AKTER, A., ISLAM, M. M., MONDAL, S. I., MAHMUD, Z., JEWEL, N. A., FERDOUS, S., AMIN, M. R., RAHMAN, M. M., 2014. Computational identification of miRNA and targets from expressed sequence tages of coffee (Coffea Arabica). *Saudi. J. Sci*., 21: 3-12.

[24] LI, X., HOU, Y., ZHANG, L., ZHANG, W., QUAN, C., CUI, Y., BIAN, S., 2014. Computational identification of conserved microRNAs and their targets from expression sequence tags of blueberry (Vaccinium corbosum). Plant Sinal. Behav., 9: e29462.

[25] GRILLO, G., TURI, A., LICCIULLI, F., MIGNONE, F., LIUNI, S., BANFI, S., GENNARINO, V. A., HORNER, D. S., PAVESI, G., PICARDI, E. AND PESOLE, G., 2010. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.*, 38: D75-80.

[26] KOZOMARA, A. AND GRIFFITHS-JONES, S., 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, 39: D152-157.

[27] HUANG, Y., NIU, B., GAO, Y., FU, L. AND LI, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, 26: 680-682.

[28] ZUKER, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31: 3406-3415.

[29] ZHANG, B., PAN, X., CANNON, C. H., COBB, G. P. AND ANDERSON, T. A., 2006. Conservation and divergence of plant microRNA genes. *Plant J.*, 46: 243-259.

[30] KRUGER, J. AND REHMSMEIER, M., 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.*, 34: W451-454.

[31] ALLEN, E., XIE, Z., GUSTAFSON, A. M. AND CARRINGTON, J. C., 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, 121: 207-221.

[32] SCHWAB, R., PPLATNIK, J. F., RIESTER, M., SCHOMMER, C., SCHMID, M. AND WEIGEL, D., 2005. Specific effects of microRNAs on the plant transcriptome. *Dev. Cell*, 8: 517-527.

[33] CONESA, A. AND GöTZ S., 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics, 2008: 1-12.

[34] CARBON, S., IRELAND, A., MUNGALL, C. J., SHU, S., MARSHALL, B. B. AND LEWIS, S., 2009. AmiGO Hub, Web Presence Working Group. AmiGO: online access to ontology and annotation data. Bioinformatics, 25: 288-289.

[35] KANEHISA, M., GOTO, S., SATO, Y., FURUMICHI, M. AND TANABE, M., 2012. KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Res.*, 40: D109-D114.

[36] ZENG, Y., 2006. Principles of micro-RNA production and maturation. *Oncogene*, 25: 6156-6162.

[37] CHANG, D. T., WANG, C. C. AND CHEN, J. W., 2008. Using a kernel density estimation based classifier to predict species-specific microRNA precursors. *BMC Bioinformatics*, 9: S2.

[38] LI, L., XU, J., YANG, D., TAN, X. AND WANG H., 2010. Computational approaches for microRNA studies: a review. *Mamm. Genome*, 21: 1-12.

[39] OBERNOSTERER, G., LEUSCHNER, P. G., ALENIUS, M. AND MARTINEZ, J., 2006 Post-transcriptional regulation of microRNA expression. *RNA*, 12: 1161-1167.

[40] HERTEL, J., LANGENBERGER, D., STADLER, P. F., 2014. Computational prediction of microRNA genes. *Methods Mol. Biol.*, 1097: 437-456.

[41] LINDIOF, A., 2003. Gene identification through large-scale EST sequence processing. *Appl. Bioinformatics*, 2: 123-129.

[42] BUZA, T., ARICK, M., WANG, H., PETERSON, D. G., 2014. Computational prediction of desease microRNAs in domestic animals. *BMC Res. Notes*, 7: 403-409.

[43] JI, Z., WANG, G., XIE, Z., ZHANG, C., WANG, J., 2012. Identification and characterization of microRNA in the dairy goat (Capra hircus) mammary gland by Solexa deep-sequencing technology. *Mol. Bio. l Rep.*, 39: 9361-9371.

[44] JI, Z., WANG, G, XIE, Z., WANG, J., ZHANG, C., DONG, F. AND CHEN, C., 2012. Identification of novel and differentially expressed micrRNAs of dairy goat mammary gland tissues using solexa sequencing and bioinformatics. *PLoS One*, 7: e49463.

[45] XU, J. H., LI, F. AND SUN, Q. F., 2008. Identification of microRNA precursors with support vector machine and string kernel. *Genom Proteom Bioinform.*, 6: 121-128.

[46] XIE, X., LU, J., KULBOKAS, E. J., GOLUB, T. R., MOOTHA, V., LINDBLAD-TOH, K., LANDER, E. S. AND KELLIS, M., 2005. Systematic discovery of regulatory motifs in human promoters and 3′-UTRs by comparison of several mammals. *Nature*, 434: 338-345.

[47] GILBERT, S. L., DOBYNS, W. B. AND LAHN, B. T., 2005. Genetic links between brain development and brain evolution. *Nat. Rev. Genet.,* 6: 581-590.

[48] OUYANG, W., RUTZ, S., CRELLIN, N. K., VALDEZ, P. A., HYMOWITZ, S. G., 2011. Regulation and functions of the IL-10 family of cytokines in inflammation and disease. *Annu. Rev. Immunol.*, 29: 71-109.

[49] LEONARDI, R., ZHANG, Y. M., LYKIDIS, A., ROCK, C. O., JACKOWSKI, S., 2007. Localization and regulation of mouse pantothenate kinase 2. *FEBS Lett.*, 581: 4639-4644.

[50] Wang, M., Wang, Q., Zhang, B., 2013. Response of miRNAs and their targets to salt and drought stresses in cotton (Gossypium hirsutum L.). *Gene*, 530: 26-32.