



Identification of Long Non-coding RNA Transcripts in *Glycyrrhiza uralensis*

Arash Rostami Azar¹, Asad Maroufi^{1,2*}

¹Department of Plant Production and Genetics, University of Kurdistan, Sanandaj, Iran

²Research Center for Medicinal Plant Breeding and Development, University of Kurdistan, Sanandaj, Iran

*Corresponding Author: Asad Maroufi, Department of Plant Production and Genetics, University of Kurdistan, Sanandaj, Iran. Tel: +98-8733620552, Fax: +98-8733620553, E-mail: a.maroufi@uok.ac.ir

Background: Chinese liquorice (*Glycyrrhiza uralensis*), an important medicinal plant, contains various valuable secondary metabolites. Secondary metabolites biosynthesis is very tightly regulated; therefore, elucidation and manipulation of the biosynthetic pathways are of great interest. Recent studies have shown that lncRNAs play important regulatory roles in many biological processes, thus identification and modification of their expression is essential to metabolic pathways for biosynthesis of secondary metabolites.

Objectives: In this study we attempted to identify non-coding RNA transcripts (lncRNAs) that may act as important regulators of diverse biological processes, including stress responses and developmental programs in *Glycyrrhiza uralensis*.

Materials and Methods: Identification of potential lncRNAs in Chinese liquorice was performed using a bioinformatics pipeline from the available EST dataset of *G. uralensis*.

Results: Bioinformatics analysis revealed that 1365 identical sequences in the range of 200 to 1286 base pair are putative lncRNAs. Only less than one percent of the predicted lncRNAs display sequence conservation with lncRNAs from other species. Moreover, 13 lncRNAs were detected as the potential precursors of 16 miRNAs. From this analysis, we also detected possible target genes of 16 known miRNA genes. The majority of the predicted miRNA target genes have important role in response to plant disease and a couple of them contribute to signalling and metabolic pathways.

Conclusion: This study demonstrates the existence of lncRNAs in *G. uralensis* which has not been found before and provides valuable resources for further understanding and characterizing of lncRNAs and also a basis for additional investigation to reveal specific roles of lncRNAs in various biological processes and particularly in response to plant diseases.

Keywords: Bioinformatics, *Glycyrrhiza uralensis*, lncRNAs, MicroRNA (MiRNA)

1. Background

Numerous protein-coding genes are involved in the development, metabolic and stress response pathways, but these genes can't sufficiently explain complexity of these biological processes in the higher organisms (1, 2). Completely sequenced genomes of eukaryotes discovered that there are far fewer protein coding genes than expected and great percentages of the transcripts are non-coding (3) little is known about lncRNAs in *Panax ginseng* C. A. Meyer, an economically significant medicinal plant species. A total of 3,688 mRNA-

like non-coding RNAs (mlncRNAs). For example, approximately 99 percent of the human genome does not encode proteins (4). This frequency is also observed in other eukaryotes such as mouse, yeast and common fruit fly, furthermore, non-coding RNAs (ncRNAs) are also abundant in plants (1). Recent studies have shown that regulatory ncRNAs have significant roles in gene regulation and are considered as the basis of an inter-gene communication system (5). Regulatory ncRNAs can be subdivided into small RNAs (sRNAs <200 nt) including microRNAs (miRNAs), small-interfering

RNAs (siRNAs), and long non-coding RNAs (lncRNAs >200 nt). LncRNAs are functional molecules with a large and varied class of transcribed RNAs. Based on position and transcriptional direction with respect to other genes, lncRNAs are classified into different subtypes such as antisense, intergenic, sense, intronic and bidirectional (6). A number of recent studies have shown that lncRNAs are involved in plant development and stress responses (3). Moreover, pathway enrichment analysis has shown that lncRNAs connected with biosynthetic and secondary metabolic pathways (7). mRNA-like non-coding RNAs (mlncRNAs) are a subset of lncRNAs that possess all the properties of mRNAs, but lack of encoding protein and can be spliced, capped and polyadenylated. mlncRNAs have been demonstrated to play important roles in several aspects of plant growth and development such as phosphate balance resulting in translational repression or site-specific cleavage. In plants, most miRNA targets are cleaved and show almost perfect complementarity with the miRNAs around the cleavage site. Here, we examined the non-protein coding gene IPS1 (INDUCED BY PHOSPHATE STARVATION1, photo morphogenesis, vernalisation, flowering regulation, cell differentiation, response to biotic and abiotic stresses and nodulation (8, 9).

It has been demonstrated that lncRNAs can be precursors of miRNAs and they are also potential targets and endogenous target mimics (eTMs) of miRNAs in plants (10). miRNAs are a class of small ncRNAs with 18-25 nt long that play important roles in gene regulation in different developmental stages (11). In plants, most miRNAs show high conservation, high expression level and similar secondary structures. miRNAs mostly target transcription factors and affect development and cell differentiation.

Chinese liquorice (*Glycyrrhiza uralensis*) is an important medicinal plant and its root contains triterpene saponins and flavonoids with various types of pharmacological properties which has become one of the hot of pharmacological studies (12). Identification of molecular factors such as lncRNA, other than known molecules such as transcription factors or encoding genes in biosynthetic and metabolic pathways, will provide important information for further application in plant improvement and metabolic engineering. Extensive researches have been done on lncRNAs in humans and animals, but such works are relatively

rare in plants and only limit to a certain model species. Recently lncRNAs and mlncRNAs in plants have gained more attention and to date, various lncRNAs and mlncRNAs have been identified in *Triticum aestivum* (2), *Panax ginseng* (3), *Medicago truncatula* (5), *Salvia miltiorrhiza* (9), *Arabidopsis thaliana* (13), *Digitalis purpurea* (14) and *Zea mays* (15). Generally, lncRNAs and mlncRNA are not well understood and attempts are required in the field of molecular genetics and bioinformatics for a more comprehensive understanding of their significance. Therefore, in this work we aimed to identify and investigate lncRNA in *G. uralensis* and finally the results are presented.

2. Objectives

Several studies have mainly focused on growth and developmental processes, secondary metabolites, discovering the protein-coding genes and gene function in *G. uralensis* (12). An increasing number of studies have also demonstrated that lncRNAs play vital roles in various biological processes and metabolic pathways in plants. Therefore, identification of lncRNAs connected to the biological processes such as stress responsive, metabolic or developmental pathways in plant can be very helpful. However, no information on lncRNAs in *G. uralensis* and their potential function are available. Thus, the use of new bioinformatics methods for prediction of lncRNAs, in *G. uralensis* can be useful. The results will increase our knowledge about the function and regulatory pattern of these important molecules. In this regard, we used bioinformatics tools to identify candidate lncRNAs, their functions, precursors of miRNAs and miRNAs potential target genes for *G. uralensis*.

3. Materials and Methods

3.1. Bioinformatics -Based Identification of LncRNAs

EST sequences were used as the source sequences for putative lncRNAs identification. These sequences were downloaded from the genomeNet database (<http://www.genome.jp/>). The dataset contained 10937 EST sequences of *G. uralensis*. Computational identification of lncRNAs was carried out according to the pipeline as illustrated in **Figure 1**. The method is generally similar to the methods for identification lncRNAs in other plant species, such as *P. ginseng* (3), *S. miltiorrhiza* (9) and *D. purpurea* (14). All 10937 EST sequences were firstly

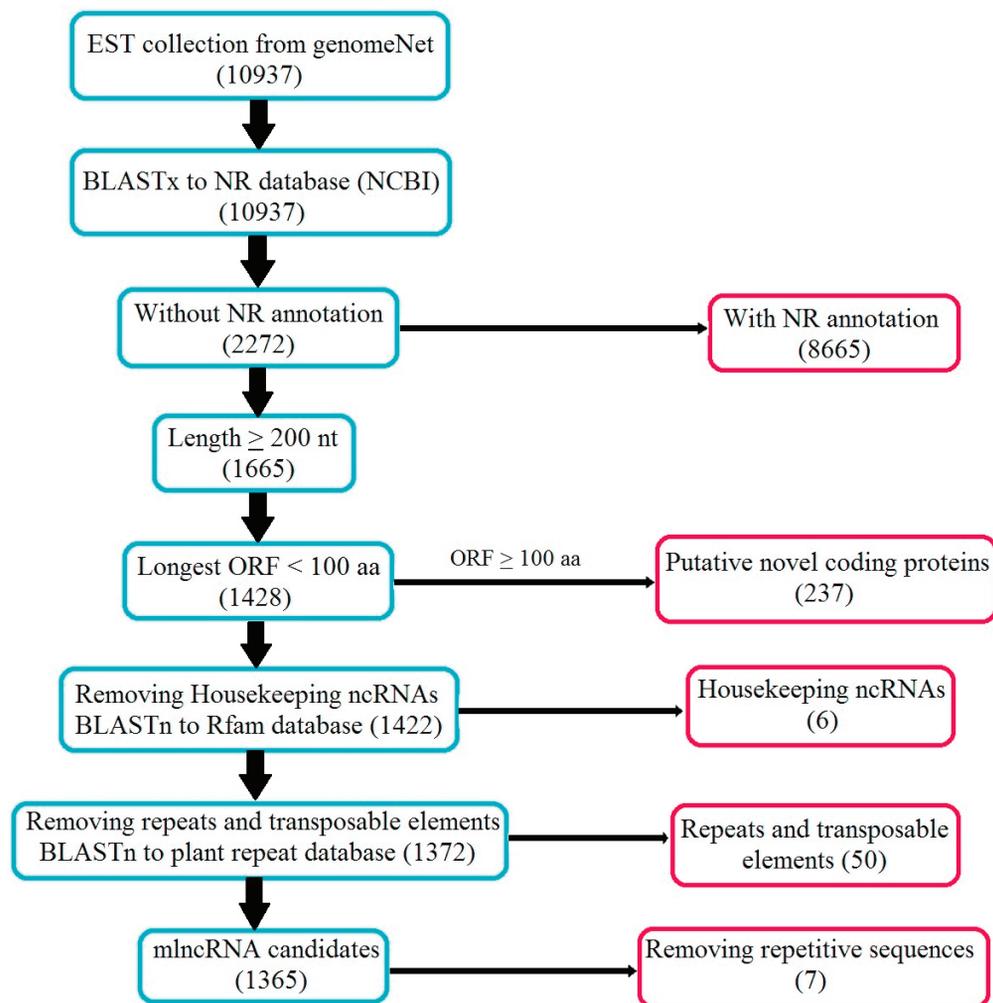


Figure 1. Framework of the steps to identify lncRNA candidates in *G. uralensis*. The number of ESTs is shown in parentheses.

aligned using translated BLAST (blastx) against NCBI non-redundant (nr) protein database ($e < 10^{-5}$) to discard the sequences with nr annotation. The remaining ESTs were filtered on the basis of their size. Then, possible open reading frames (ORFs) for the ESTs without annotation were examined using OrfFinder (<https://www.ncbi.nlm.nih.gov/orffinder/>) ORFs with a cut-off of 100 amino acids to distinguish non-protein-coding RNAs (npcRNAs) from protein-coding transcripts (13). Next, the housekeeping npcRNAs were identified by searching against Rfam 12.1 database (including 2474 families), using nucleotide BLAST (BLASTn). The remaining sequences were aligned using BLASTn ($e < 10^{-5}$) against plant repeat database to eliminate the repeats and transposable elements. The maintained ESTs were finally regarded as putative lncRNAs.

3.2. Classification of LncRNAs Based on Sequence Similarity

Sequence homology of candidate lncRNAs was assessed using BLASTn with a cut-off of $e < 10^{-5}$. Next, the level of sequence conservation of the putative identified lncRNAs by searching against known ncRNAs in NONCODE2016 database was inspected using BLASTn with e-value cut-off of 10^{-5} .

3.3. Identification of LncRNAs Precursors of miRNA's

Plant known miRNA sequences from miRBase (release 21) (<http://www.mirbase.org/>) and PMRD (April 2016) (<http://bioinformatics.cau.edu.cn/PMRD/>) databases were used to find miRNA precursors within *G. uralensis* candidate lncRNAs. lncRNAs were searched against plant miRNA sequences allowing up to three mismatches

using standalone BLAST +2.2.26. The e-value of 10^{-2} and word size of 4 were selected as cut-off to this BLAST. Next, the secondary structures of lncRNAs with maximum three mismatches were predicted on the 'mfold web server' (<http://unafold.rna.albany.edu/?q=mfold>) 'mfold web server', describes a number of closely related software applications available on the World Wide Web (WWW). The secondary structure with the lowest MFEI (minimal folding free energy index) was selected for further analysis. Selected structures were then manually checked with the criteria described from previous studies (16, 17). The substantial criteria were as follows: (a) appropriate precursor secondary structure with MFEI less than minus 30 Kcal.mol⁻¹, (b) miRNA in single arm with no nick or loop, (c) four or fewer mismatches in base-pairing between the miRNA and its opposite arm, called miRNA*, (d) A+U amount between 30 to 75 percent, and two or fewer asymmetric humps between miRNA/miRNA*.

3.4. Prediction of miRNA Target Genes

'PsRNATarget web server' (<http://plantgrn.noble.org/psRNATarget/>) was utilized for prediction of identified miRNAs targets. Default parameters were set as input for the analysis but only maximum expectation value was adjusted to 2. *Medicago truncatula* unigenes "DFCI Gene index (MTGI) version 11, released on March 23, 2011" were selected as preloaded transcript/genomic library for target search. Predicted target genes were checked with the following criteria (17): first, up to 4 mismatches between identified miRNAs and their putative targets, second, no mismatches in cleavage site, meaning base-pairing at 10th and 11th position, third, not more than one mismatch between 2nd and 12th position with maximum three mismatches between positions 12th and 25th, and finally not more than two consecutive mismatches. The identified target genes were blasted against non-redundant protein sequences with an e-value of 10^{-5} . UniProt (release August 2016; <http://www.uniprot.org/>) and InterPro 59.0 (<https://www.ebi.ac.uk/interpro/>) databases were used to find out the function of identified proteins.

3.5. RT-PCR

Differentially expression analysis was performed for two randomly selected lncRNA transcripts. Total RNAs were extracted from root tissue of plant samples of *Glycyrrhiza glabra* and *Glycyrrhiza uralensis*. First-

strand cDNAs were synthesized using Moloney Murine Leukemia Virus (M-MuLV) Reverse Transcriptase (Vivantis), according to the manufacturer's instructions. RT-PCR was performed using specific primers for lncRNAs and reference gene (Actin). Actin is selected as an internal standard (reference gene) for calculation of relative gene expression levels. The primers were designed using Primer3Plus online software (**Table 1**). The PCR products were loaded on 1.2% agarose gels in TBE buffer and stained using Ethidium bromide. An image of the RT-PCR on agarose gel was acquired. The intensity of the band corresponding to the reference and target genes was obtained using GelQuant.NET software (V 1.7.8; <http://biochemlabsolutions.com/GelQuantNET.html>). Band intensity reflects the number of copies of the gene transcript. The intensity of the lncRNAs band (target gene) was finally divided by the intensity of the reference gene. The ratio of target gene to reference is used to compare target transcript abundance in different plant samples.

4. Results

4.1. Prediction of lncRNAs in *G. uralensis*

In the first step 10937 ESTs of *G. uralensis* were investigated with BLASTx against non-redundant protein database and 2272 (20.77%) ESTs were found that had no significant similarity, therefore these sequences were considered as ncRNAs. Amongst 2272 ESTs, 1665 transcripts larger than 200 nt in length were selected for further examination. The 1665 ESTs contained 237 putative encoding novel proteins with viable ORF (more than 100 amino acids; Supplementary S1) and 6 new housekeeping ncRNAs (**Table 2**). After discarding potential repeats and transposable elements, 1365 ESTs were finally identified as functional lncRNA candidates (Supplementary S2). The nominated lncRNAs are about 12.5% of total available EST sequences of *G. uralensis*. This was in line with previous reports, 11.75% for *P. ginseng* (3), 11.3% for *D. purpurea* (14) and 12.25% for *S. miltiorrhiza* (9). However, the ratio for *Zea mays* was lower 5.41% (15).

Further analysis showed that most of the identified lncRNAs (94.5%) vary amongst 200 and 600 nucleotides with an average of 393 nt (**Fig. 2**). The longest lncRNA has 1286 nt and the shortest one has 200 nt. These findings were concordant with observational analyses in other plant species; *P. ginseng* (3), *S. miltiorrhiza* (9) and *D. purpurea* (14).

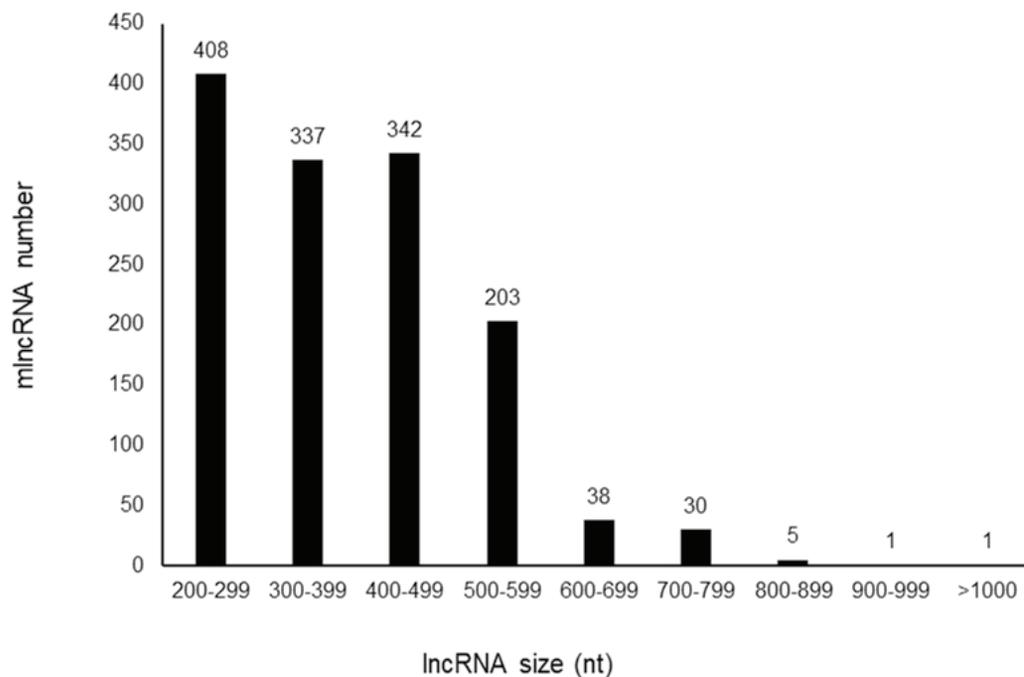
Table 1. Primers used for RT-PCR

gene	primer sequence	size
Actin-F	TTGGGATGGGTCAAAAGG	398
Actin-R	ACGAAGGATGGCATGAGG	
T20098:699-F	GGCTGGTCCTGGATGTAAAG	480
T20098:699-R	GCACTCATCTAAGAGGCAACAAC	
T20098:2650-F	TTTCTGTAGCCTCTTCCCATC	755
T20098:2650-R	TGTTGTTGTCGTCGTCTTCTC	

F, Forward; R, reverse

Table 2. New candidate housekeeping npcRNAs for *G. uralensis*

npcRNA type in Rfam	npcRNA length (nt)	sequence id
snoR38	207	T20098:104 (N)
enod40	716	T20098:3972 (N)
snoR83	337	T20098:5325 (N)
MIR396	509	T20098:6122 (N)
snoZ267	413	T20098:8476 (N)
LSU rRNA eukaryal	271	T20098:10472 (N)

**Figure 2.** Size distribution of candidate lncRNAs in *G. uralensis*.

Among the predicted lncRNAs, 38 (2.78%) are in the range of 600-700 nt and 37 (2.71%) contain more than 700 nt. The longer predicted lncRNA, with the size over 600 bp, is very likely to be true lncRNA (14). However, some short lncRNAs are possibly UTR fragments of protein coding transcripts. Hence, more experimental studies are needed to identify those sequences.

4.2. lncRNAs Classification Based on Sequence Similarity

lncRNA classification based on sequence similarity assists in the prediction of their structure and function. Moreover, the existence of homology among lncRNAs may indicate the existence of shared ancestry. Using BLASTn with an e-value of 10^{-5} , no sequence homology was found among 1222 predicted lncRNAs for *G. uralensis* and they are likely to be single copy. Other 143 lncRNAs (10.48%) were grouped into 60 classes (Supplementary S3). Moreover, to explore lncRNAs conservation across species, alignment against NONCODE database was performed using BLASTn ($e < 10^{-5}$). Only 11 lncRNAs from 1365 putative lncRNAs in *G. uralensis* showed sequence conservation with lncRNAs in other species (Supplementary S4). lncRNAs appear poorly conserved in higher plant species (18). Low conservation of lncRNAs has been reported for other organisms, such as *F. vesca* (18), *M. truncatula* (5), *D. purpurea* (14), *P. ginseng* (3), *Arabidopsis thaliana* (13), *T. aestivum* (2), *D. melanogaster* (19) and *Mus musculus* (20).

4.3. Identification of lncRNAs Precursors of miRNA's

Some lncRNAs can also serve as the precursor of miRNAs (21). miRNAs are small ncRNAs that cover important functions in organ development, cell differentiation, different levels of post-transcriptional regulation and defence responses (11). High sequence conservation among organisms is one of the main miRNA characteristics (16, 22). Therefore, the conservation is a key factor to find miRNAs using the bioinformatics analysis. Accordingly, we identified 13 lncRNAs as the precursors of 16 miRNAs (Table 3). Four lncRNAs as the precursors of miRNAs have length ranging from 200 nt to 399 nt, seven lncRNAs from 400 nt to 599 nt and two lncRNAs with the length over 600 nt. The majority of these predicted miRNA precursors have the length over 300 nt. The predicted miRNAs were classified into 11 families (Table 3). Among them, *miR482* has 3 and *miR414*, *miR5021* and *miRf11142* have 2 members.

Seven families including *miR23*, *miR156*, *miR3946*, *miR5213*, *miR6169*, *miRf10014* and *miRf10465* have only one member. Nearby 0.95% of total predicted lncRNAs or 0.11% of total *G. uralensis* ESTs were recognized as the precursors of miRNAs. This ratio (0.11%, in relation to total ESTs) is higher than the ratios reported for other plant species with an average of 0.010% (16); 0.023% for *Aegilops speltoides*, 0.06% for *D. purpurea*, 0.006% for *A. thaliana*, 0.005% for *M. truncatula* and 0.009% for *Z. mays*. Furthermore, secondary structures of the candidate lncRNAs with miRNA region were drawn (Supplementary S5). These structures exhibit miRNA precursors of plant species.

4.4. Prediction of Identified miRNA Target Genes

The psRNATarget web server was used to find miRNA target prediction. It works based on high degree of complementarity of miRNAs with their target genes. Using psRNATarget, 37 target genes were screened out for 4 miRNAs (Supplementary S6). These miRNAs target variable number of genes with maximum 22 genes by miR3946. Moreover, no target gene was discovered for 7 miRNAs. Additionally, functions of the potential target genes were determined via UniProt and InterPro databases. Most of the potential miRNA target genes are involved in defence responses, although some are transcription factors, and a few of them contribute to signal transduction and metabolic processes.

4.5. RT-PCR

Gene expression at transcript level of two candidate lncRNA (T20098:699 and T20098:2650) was studied. First of all, the lncRNA transcripts were detected in all *G. uralensis* plant root tissue (Fig. 3) but different expression level has been observed for both genes in the different plants (Fig. 3). The expression level of T20098:699 was higher than the level of T20098:2650 lncRNA. Furthermore, the transcript levels of these two lncRNAs were also studied in *G. glabra* plants. Only the transcript of T20098:699 were detected in root tissue of *G. glabra*.

5. Discussion

G. uralensis, known as Chinese liquorice, produces a variety of phytochemicals and secondary metabolites such as saponins, chalcones, flavonoids, and triterpenes. Important saponin like glycyrrhizic acid from root extract have medicinal properties such as protection against hepatotoxicity and anti-inflammatory (24).

Table 3. The predicted miRNAs for *G. uralensis*

miRNA Family	Candidate miRNA (<i>G. uralensis</i>)	Mature microRNA	ME	Sequence id	SP	EP	e-value	% A+U	MFE
miR482	gu-miR482	TCTTCCAATTCGCCCATTCCTA	24/24	T20098:1490 (N)	212	235	1.00E-07	50	-154.25
	gu-miR482c	GTTCCATATTCCTCCCATGCCAC	19/22	T20098:2291 (N)	124	145	7.00E-04	45.45	-117.26
	gu-miR482c-3p	TCCCAATTCGCCCATTCCT	21/21	T20098:1490 (N)	214	234	4.00E-06	47.61	-154.25
miR414c	gu-miR414c	TCATCATCATCATCT	15/15	T20098:2522 (N)	286	300	3.00E-03	66.66	-157.36
	gu-miR414c	TCATCATCATCATCT	15/15	T20098:5335 (N)	52	66	3.00E-03	66.66	-64.46
miR5021	gu-miR5021	TGAGAAGAAGAAGAAGATC	18/20	T20098:6795 (N)	91	110	1.00E-04	65	-61.41
	gu-miR5021	TGAGAAGAAGAAGAAGAAGC	18/20	T20098:9257 (N)	201	220	1.00E-04	60	-107.24
miRf11142	gu-miRf11142-akr1	GTATAACATCATGAGCAGTCA	19/21	T20098:8886 (N)	146	166	5.00E-05	61.9	-106.86
	gu-miRf11142-akr2	CGATAGCATCATGAGCAATCA	19/21	T20098:8886 (N)	192	212	5.00E-05	57.14	-106.86
miR156	gu-miR156z	GTTGGAGTGAAGGGAGAG	15/18	T20098:204 (N)	143	160	4.00E-03	44.44	-155.4
miR23	gu-miR23-npr	TCCCAATTCGCCCATTCCTA	20/22	T20098:1490 (N)	214	235	2.00E-03	50	-154.25
miR6169	gu-miR6169	TCCTATTCTCTTCTCTCTCT	18/22	T20098:7456 (N)	54	75	8.00E-03	66.66	-75.08
miRf10014	gu-miRf10014-akr	TAGTGCACATGGTGTGT	17/20	T20098:8342 (N)	201	220	7.00E-03	50	-79.73
miRf10465	gu-miRf10465-akr	TGAGCCTCTGCACCTCCGTC	20/22	T20098:1832 (N)	515	536	9.00E-03	40.9	-163.04
miR5213	gu-miR5213	TACGGGTGTCTCACCTCTGAG	20/22	T20098:1896 (N)	101	122	2.00E-04	45.45	-136.13
miR3946	gu-miR3946	TTGTAGAGAGAGAGAGAGGCAC	22/24	T20098:2137 (N)	21	44	1.00E-02	50	-91.76

SP: Start Point, EP: End Point, ME: Match Extent, MFE: Minimal Free Energy



Figure 3. RT-PCR of lncRNAs (T20098:699, T20098:2650 and Actin) for 4 different plants. M, DNA ladder; numbers indicate different plants.

Terpenoids and flavonoids also have beneficial effects on human health (14). Discovering the genes and their regulatory networks in a given biosynthetic pathway in *G. uralensis* is the main objective of metabolic engineering in liquorice plants.

lncRNAs contribute to gene regulation through a variety of mechanisms. To date, many lncRNAs

with several actions have been characterized. The main functions for lncRNAs are defined as follows: signal, decoy, scaffold, guide, enhancer RNAs, and short peptides (25). Besides, lncRNAs can also act as siRNAs and miRNA precursors in plants (2, 14, 15) and animals (20). Moreover, the effect of regulatory ncRNAs on plant traits including yield, nutrition,

phosphate metabolism, and response to biotic and abiotic stresses, have been demonstrated. lncRNAs can function as miRNA target mimics as well (8). There is also a complex relationship between lncRNAs and miRNAs in plants. Therefore, discovery of lncRNAs and their target genes in *G. uralensis* can be valuable to further investigate their roles which might provide new indications to elucidate how lncRNAs and their targets play role in plant developmental processes and metabolic pathways. Using bioinformatics approaches a total of 1365 putative lncRNAs were identified from *G. uralensis* EST database (10937 sequences). This is about 12.5% of the available ESTs. This proportion is in the range of ratios were obtained for other plant species which have yet been studied. According to the available data so far, about 10-12% of plant transcripts are possibly lncRNAs. Among the entire lncRNAs candidate for *G. uralensis*, 75 identified lncRNAs were longer than 600 bp which in *D. purpurea* 52 lncRNAs (14) and in *P. ginseng* 499 lncRNAs (3) were also longer than 600 bp. 1222 of the identified lncRNAs for *G. uralensis* (89.5%) did not share significant sequence similarity. However, 143 lncRNAs (10.48%) show similarity and are grouped into 60 classes. In *D. purpurea* 320 lncRNAs were also grouped into 140 classes (14), in *P. ginseng* 942 lncRNAs were classified in 245 families (3) and in *S. miltiorrhiza* 2030 lncRNAs were classified into 470 families (9). Nevertheless, larger lncRNAs classes were identified in *S. miltiorrhiza* (9) and *P. ginseng* (3) including 530 and 344 members respectively. Fewer degrees of similarity of lncRNAs in plant species are probably due to the lower complexity of lncRNAs. Moreover, only 11 identified lncRNAs (0.8% of total identified lncRNAs) in *G. uralensis* displayed sequence conservation across different species such as mouse, arabidopsis and human. These results are consistent with findings in other species, for example, only 8 from 2660 lncRNAs in *D. purpurea* (14), 32 from 5444 in *S. miltiorrhiza* (9) and 20 from 3688 in *P. ginseng* (3) showed sequence conservation. These statistics suggest that lncRNAs have less conserved motifs than protein-coding genes.

Genome analysis have revealed that a significant portion of long non-coding transcripts in eukaryotes act as precursors of miRNAs (26). Therefore, we predicted the potential miRNAs which can be produced by lncRNAs. Alignment of candidate lncRNAs for *G. uralensis* against the miRBase resulted in the

identification of 13 miRNA- carrying lncRNAs. These lncRNAs were found to be the precursor of 16 possible novel miRNAs for *G. uralensis*. Many mature miRNAs are evolutionarily conserved from species to species within the same kingdom (27). miRNA encoding genes in one species may arise as orthologues or homologs in other species (28). Thus, it was computationally possible to predict new miRNA- carrying lncRNAs. The longest precursor lncRNAs was 806 nt and the shortest one was 276 nt. These identified miRNAs were in the range of 15 to 25 nt. This result is in accordance with the records obtained for other plant species such as arabidopsis (29), tobacco (30) helianthus (31) tomato (32) and potato (33). Computational analysis revealed that only 0.11% of total ESTs of *G. uralensis* detected as miRNA precursors. This proportion is at a higher level reported for other plant species with an average of 0.010% (16). Altogether these statistics indicate the potential involvement of lncRNAs and their predicted miRNA in the regulation of biological processes in *G. uralensis*.

To explore the role of miRNAs in the regulation of gene expression, therefore, it is required to find their target genes. Via psRNATarget server the miRNA target genes were predicted. Most of the predicted target genes belong to the plant defence responses; however other target genes such as transcription factors, signal transduction and genes involving in metabolism were identified. Furthermore, the majority of the defence response genes were disease resistance genes encoding proteins that contain leucine-rich repeat domain (LRRs). In plants, disease resistance genes commonly encode proteins with nucleotide-binding site leucine-rich repeat (NBS-LRR) (34). The increasing data indicate that small RNAs including miRNAs are involved in the regulation of plant immunity (35). In wheat 125 putative stress responsive lncRNAs were identified (2). Additionally, 4 lncRNAs which participate in defence against stripe rust pathogen have been isolated from ESTs of wheat (36). In foxtail millet 584 drought-responsive lncRNAs have been identified (37). Consequently, plants, in response to pathogens and stresses in the absence of adaptive immunity system, have likely developed a mechanism using lncRNAs which participate in stress response mechanism. Therefore, lncRNAs should be considered more in plant stress responses in the future studies. Among the pool of predicted target genes of miRNAs in *G. uralensis*, some were transcription

factors. Transcription factors were also found to be targeted by miRNAs in plants (38). In *Camellia sinensis* 20% of the identified potential target genes encode transcription factors (39). Transcription factors like; APETALA2, WRKY3, DELLA, MYB were also detected in *Artemisia annua* L. as the targets of miRNAs (31). Moreover, a number of miRNAs of *Vigna unguiculata* L. were detected to target transcription factors and protein with function in the process of signal transduction (40). Hence, recognized lncRNAs as the miRNA precursors which target transcription factors will help plant scientists in understanding the mechanism of action and interaction between these regulators in the plant growth and development.

RT-PCR for two lncRNA in a number of root samples was carried out to verify transcript expression patterns and how abundant the transcripts. Our study confirmed the differentially expression of lncRNAs in *G. uralensis* plant roots. Moreover, the transcript levels of lncRNAs were examined in *G. glabra* another liquorice species and only for one lncRNA in root tissue transcript was detected. This issue may be related to the poor evolutionary conservation of the lncRNAs (23). Additionally, the lncRNAs are usually expressed in a more tissue-specific manner with lower level than mRNAs of protein-coding genes (1).

6. Conclusions

This is the first step towards the identification of lncRNAs and miRNA- carrying lncRNAs in *G. uralensis* through *in silico* studies. In summary, a total of 1365 putative lncRNAs have been identified from available EST sequences of *G. uralensis*. All these lncRNAs and their miRNA target genes are reported for the first time. The predicted target genes are involved in different biological processes and pathways with molecular functions mainly as plant defence genes, transcription factor, signal transduction and metabolic genes. Identification of lncRNAs, miRNAs and their target genes provide new information for understanding the functions of lncRNAs in *G. uralensis*. Moreover, characterization of lncRNAs and their functions in different biological processes, especially plant development, disease resistance and stress responses for this important medicinal plant would be very valuable.

Acknowledgements

The authors would like to thank Dr. Jafar Afshinfar for help with editing the manuscript.

References

1. Statello L, Guo CJ, Chen LL, *et al.* Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol.* 2021;**22**:96-118. doi.org/10.1038/s41580-020-00315-9
2. Xin M, Wang Y, Yao Y, Song N, Hu Z, Qin D, Xie C, Peng H, Ni Z, Sun Q. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol.* 2011;**11**:1-13. doi:10.1186/1471-2229-11-61
3. Wang M, Wu B, Chen C, Lu S. Identification of mRNA-like non-coding RNAs and validation of a mighty one named MAR in *Panax ginseng*. *J Integr Plant Biol.* 2015;**57**(3):256-270. doi:10.1111/jipb.12239
4. Lee H, Zhang Z, Krause HM. Long noncoding RNAs and repetitive elements: junk or intimate evolutionary partners. *Trends Genet.* 2019;**35**:892-902.
5. Wen J, Parker BJ, Weiller GF. In Silico identification and characterization of mRNA-like noncoding transcripts in *Medicago truncatula*. *In Silico Biol.* 2007;**7**:485-505.
6. Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. *Nat Struct Mol Biol.* 2015;**22**(1):5-7.
7. Bhatia G, Sharma S, Upadhyay SK, Singh K. Long Non-coding RNAs Coordinate Developmental Transitions and Other Key Biological Processes in Grapevine. *Sci Rep.* . 2019;**9**(1):1-14. doi: 10.1038/s41598-019-38989-7
8. Quan M, Chen J, Zhang D. Exploring the secrets of long noncoding RNAs. *Int J Mol Sci.* 2015;**16**:5467-5496. doi:10.3390/ijms16035467
9. Li D, Shao F, Lu S. Identification and characterization of mRNA-like noncoding RNAs in *Salvia miltiorrhiza*. *Planta.* 2015;**241**:1131-1143. doi:10.1007/s00425-015-2246-z
10. Angrand PO, Vennin C, Le Bourhis X, Adriaenssens E. The role of long non-coding RNAs in genome formatting and expression. *Front Genet.* 2015;**6**:1-12. doi:10.3389/fgene.2015.00165
11. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet.* 2004;**5**:522-531. doi:10.1038/nrg1415
12. Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K, Muranaka T. Licorice -amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proc Natl Acad Sci USA.* 2008;**105**(37):14204-14209. doi:10.1073/pnas.0803876105
13. MacIntosh GC, Wilkerson C, Green PJ. Identification and Analysis of Arabidopsis Expressed Sequence Tags Characteristic of Non-Coding RNAs. *Plant Physiol.* 2001;**127**:765-776. doi: 10.1104/pp.010501
14. Wu B, Li Y, Yan H, Ma Y, Luo H, Yuan L, Chen S, Lu S. Comprehensive transcriptome analysis reveals novel genes involved in cardiac glycoside biosynthesis and mlncRNAs associated with secondary metabolism and stress response in *Digitalis purpurea*. *BMC Genomics.* 2012;**13**:15. doi:10.1186/1471-2164-13-15
15. Boerner S, McGinnis KM. Computational identification and functional predictions of long noncoding RNA in *Zea mays*. *PLoS One.* 2012;**7**(8):e43047. doi: 10.1371/journal.pone.0043047
16. Zhang B, Pan X, Cobb GP, Anderson TA. Identification and

- characterization of new plant microRNAs using EST analysis. *Cell Res.* 2005;**15**:336-360. doi:10.1038/sj.cr.7290302
17. Patanun O, Lertpanyasampatha M, Sojikul P, Viboonjun U, Narangajavana J. Computational identification of MicroRNAs and their targets in cassava (*Manihot esculenta* Crantz.). *Mol Biotechnol.* 2013;**53**(3):257-269. doi:10.1007/s12033-012-9521-z
 18. Kang CY, Liu ZC (2015) Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. *BMC Genomics.* 2015;**16**(815). doi: 10.1186/s12864-015-2014-2
 19. Inagaki S, Numata K, Kondo T, Tomita M, Yasuda K, Kanai A, Kageyama Y. Identification and expression analysis of putative mRNA-like non-coding RNA in *Drosophila*. *Genes to Cells.* 2005;**10**:1163-1173. doi:10.1111/j.1365-2443.2005.00910.x
 20. Numata K1, Kanai A, Saito R, et al. Identification of Putative Noncoding RNAs Among the RIKEN Mouse Full-Length cDNA Collection. *Genome Res.* 2003;**13**(6b):1301-1306. doi:10.1101/gr.1011603
 21. Fan G, Cao Y, Wang Z (2018) Regulation of Long Noncoding RNAs Responsive to Phytoplasma Infection in *Paulownia tomentosa*. *Int J Genomics.* doi: 10.1155/2018/3174352
 22. Li L, Xu J, Yang D, Tan X, Wang H. Computational approaches for microRNA studies: a review. *Mamm Genome.* 2010;**21**:1-12. doi: 10.1007/s00335-009-9241-2
 23. Necsulea A, Kaessmann H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat Rev Genet.* 2014;**15**:734-748. doi: 10.1038/nrg3802
 24. Shibata S. A Drug over the Millennia: Pharmacognosy, Chemistry, and Pharmacology of Licorice. *Yakugaku Zasshi.* 2000;**120**(10):849-862. doi:10.1248/yakushi1947.120.10_849
 25. Li X, Wu Z, Fu X, Han W. LncRNAs: insights into their function and mechanics in underlying disorders. *Mutat Res Rev Mut Res.* 2014;**762**:1-21. doi: 10.1016/j.mrrev.2014.04.002
 26. Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua NH. Genome-wide analysis uncovers regulation of long intergenic non-coding RNAs in *Arabidopsis*. *Plant Cell.* 2012;**24**(11):4333-4345. doi: 10.1105/tpc.112.102855
 27. Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA. Conservation and divergence of plant microRNA genes. *Plant J.* 2006;**46**(2):243-259. doi: 10.1111/j.1365-313X.2006.02697.x
 28. Weber MJ. New human and mouse microRNA genes found by homology search. *FEBS J.* 2005;**272**(1):59-73. doi: 10.1111/j.1432-1033.2004.04389.x
 29. Wang XJ, Reyes JL, Chua NH, Gaasterland T. Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. *Genome Biol.* 2004;**5**(9):R65. doi:10.1186/gb-2004-5-9-r65
 30. Frazier TP, Xie F, Freistaedter A, Burklew CE, Zhang B. Identification and characterization of microRNAs and their target genes in tobacco (*nicotiana tabacum*). *Planta.* 2010;**232**:1289-1308. doi:10.1007/s00425-010-1255-1
 31. Barozai MYK. Identification of microRNAs and their targets in *Artemisia Annu* L. *Pak J Bot.* 2013;**45**(2):461-465.
 32. Din M, Barozai MYK. Profiling microRNAs and their targets in an important fleshy fruit: Tomato (*Solanum lycopersicum*). *Gene.* 2014;**535**:198-203. doi:10.1016/j.gene.2013.11.034
 33. Din M, Barozai MYK, Baloch IA. Identification and functional analysis of new conserved microRNAs and their targets in potato (*Solanum tuberosum* L.). *Turk J Botany.* 2014;**38**:1199-1213. doi:10.3906/bot-1405-105
 34. McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 2006;**7**(4):212.
 35. Fei Q, Zhan Y, Xia R, Meyers BC. Small RNAs add zing to the zig-zag-zig model of plant defenses. *Mol Plant Microbe Interact.* 2016;**29**:165-169. doi: 10.1094/MPMI-09-15-0212-FI
 36. Zhang H, Chen X, Wang C, Xu Z, Wang Y, Liu X, Kang Z, Ji W. Long non-coding genes implicated in response to stripe rust pathogen stress in wheat (*Triticum aestivum* L.). *Mol Bio Rep.* 2013;**40**(11):6245-6253. doi: 10.1007/s11033-013-2736-7
 37. Qi X, Xie S, Liu Y, Yi F, Yu J. Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. *Plant Mol Biol.* 2013;**83**(4-5):459-473. doi: 10.1007/s11103-013-0104-6
 38. Barozai MYK, Din M, Baloch IA. Structural and functional based identification of the bean (*Phaseolus*) microRNAs and their targets from expressed sequence tags. *J Struct Funct Genomics.* 2013;**14**:11-18. doi: 10.1007/s10969-013-9152-z
 39. Prabu GR, Mandal AK. Computational identification of miRNAs and their target genes from expressed sequence tags of tea (*Camellia sinensis*). *Genom Proteomics Bioinformatics.* 2010;**8**(2):113-121. doi: 10.1016/S1672-0229(10)60012-5
 40. Gul Z, Barozai MYK, Din M. In-silico based identification and functional analyses of miRNAs and their targets in Cowpea (*Vigna unguiculata* L.). *AIMS Genet.* 2017;**4**(2):138-165. doi: 10.3934/genet.2017.2.138