J. Agr. Sci. Tech. (2022) Vol. 24(6): 1457-1471

Identification of miRNAs and Their Target Genes in *Taraxacum* spp

A. A. Karimi[1], M. R. Naghavi[1*], S. A. Peighambari[1], A. Sobhani[2], and A. Rasoulnia[1]

ABSTRACT

MicroRNAs are endogenous noncoding RNA that play vital roles in all plant cellular metabolic processes by mediating target gene expression. To date, miRNAs in *Taraxacum* spp., which is an important industrial plant, have remained largely unknown. In the present study, 970 miRNAs from 399 families were identified in *Taraxacum spp* by conducting computational approaches. The most frequent miRNAs in *Taraxacum* spp was miR5021. According to the KEGG results, miR5021, miR838, and miR1533 are related to the terpene biosynthesis pathway, while miR5015b, and miR1436 are involved in the starch and sucrose biosynthesis pathways. Quantitative real-time PCR assay was performed to validate the expression levels of five predicted miRNAs and 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase (HMGCR) and invertase as the target genes. Results indicated that the highest relative expression of miR1533 and miR1436 occurred in the flower, while the highest transcripts levels of miR5015b were observed in the stem. In addition, the higher relative expression level of the miR5021 and miR838 was consistent with the lower expression level of the *HMGCR* gene in all tissue, suggesting that miR5021 and miR838 are involved in regulating *HMGCR* gene expression. Since mevalonate pathway is the main source of isopentenyl pyrophosphate, which is used in the synthesis of rubber, miR5021 and miR838 play an important role in the production of rubber by regulating the expression of HMGCR enzyme. These findings will accelerate future perspective studies on the regulatory mechanisms of miRNAs in *Taraxacum kok-saghyz*.

Keywords: MicroRNA, Natural rubber, RNA-seq, *Taraxacum kok-saghyz*, Terpene.

INTRODUCTION

Taraxacum, as a diploid herbaceous plant, belongs to a large genus of the *Asteraceae* family. *Taraxacum kok-saghyz* and *Taraxacum officinale* are two common species of *Taraxacum* genus (Nowicki *et al*., 2019). The highest amount of latex is present in the root of *T. officinale*, which is composed of Triterpene Acetates (TritAc), Phenolic Inositol Esters (PIEs), and the sesquiterpene lactone Taraxinic acid b-D-Glucopyranosyl ester (TA-G) (Huber *et al*., 2015). The root of *T. kok-saghyz* could produce a large amount of NR (up to ~20% dry weight) (Van Beilen and Poirier, 2007). Due to the rubber-producing properties of *T. kok-saghyz*, this plant can be a substitute for the traditional source of Natural Rubber (NR) produced by *Hevea brasiliensis* (Ramirez-Cadavid *et al*., 2017; Karimi *et al*., 2021). NR is Cis-structure biopolymers of isoprene units (Hayashi, 2009) and is used extensively in many applications and products including tires, medical devices, surgical gloves, and various engineering and consumer products (Mooibroek and Cornish, 2000). In addition to rubber, this plant has another important secondary metabolite

[1] Division of Biotechnology, Department of Agronomy and Plant Breeding, College of Agricultural and Natural Resources, University of Tehran, P.O. Box: 31587–11167, Karaj, Islamic Republic of Iran.
[2] Agricultural Biotechnology Research Institute of Iran - Isfahan Branch, Agricultural Research, Education and Extension Organization (AREEO), P. O. Box: 85135-487, Isfahan, Islamic Republic of Iran.
*Corresponding author; e-mail: mnaghavi@ut.ac.ir

[DOI: 10.52547/jast.24.6.1457]

named inulin, which plays an important role in human health as a probiotic compound (Patel and Goyal, 2012).

Unfortunately, in recent years, production of NR in the *H. brasiliensis* has been challenged with some threat such as high demand due to the economic development of some countries, leaf blot, and white root rot disease caused by fungal pathogens (Luo *et al*., 2017; Salehi *et al*., 2022). Therefore, attempts have been launched to develop *T. kok-saghyz* as an alternative source of NR production.

Briefly, natural rubber consists of polymers of the organic compound isoprene, which is synthesized from the allylic Pyrophosphate (allylic-PP) molecule and the monomer Isopentenyl Pyrophosphate (IPP) via 2-C-1-Deoxy-D-Xylulose-5-Phosphate/Methyl-D-Erythritol-4-Phosphate (DOXP/MEP) and Mevalonate (MVA) pathways. The main source of IPP used in the formation of rubber chains is the MVA pathway, and the enzyme HMGCR is a key regulator of this pathway (Panara *et al*., 2018). Also, it is well proven that rubber production in the root of *T. kok-saghyz* could negatively be affected by the main storage carbohydrate, inulin (Luo *et al*., 2017).

Sucrose is the end product of photosynthesis and the primary sugar transported in the phloem of most plants. One of the important enzymes of sucrose metabolism is invertase, which hydrolyzes sucrose to glucose and fructose monosaccharides. (Wang *et al*., 2017). Since the most important secondary metabolite of *T. kok-saghyz* is rubber and various regulatory factors such as miRNAs affect the expression of genes involved in the rubber pathway, the study of miRNAs regulating this pathway is extremely valuable.

MicroRNAs are key and effective biological elements in regulating gene networks in plants (Samad *et al*., 2017). These biological elements are a group of non-coding RNAs that regulate gene expression at different levels of transcription, processing, RNA stability, and translation through the process of RNA interference. MiRNAs play a critical role in various aspects of plant development and response to stress by regulation of gene expression at the transcriptional or post-transcriptional level (Sabzehzari and Naghavi, 2019). Besides, numerous miRNAs and their target genes have been identified in various plant species such as *Malus domestica*, *Gossypium arboreum*, *Panicum virgatum*, *Nicotiana tabacum*, etc. (Frazier *et al*., 2010; Xie *et al*., 2010; Wang *et al*., 2012). In general, three methods of direct cloning, direct genetics, and computational bioinformatics methods are used to identify miRNAs, while three methods of bioinformatics, biochemistry, and omics-based methods are used to recognize their target genes (Li *et al*., 2010). In the meantime, computational methods are faster and much less expensive than other methods and are highly used to identify potential conserved miRNAs. Computational methods, based on homology and comparing the database of identified miRNAs (http://www.mirbase.org) with databases of the target organism (such as transcriptome database), are used to identify potentially conserved miRNAs (Li *et al*., 2010; Tian *et al*., 2015).

In this study, the EST database and paired-end short reads of the *T. kok-saghyz* and *T. officinal*e were used for the computational prediction of miRNAs and their targets. Moreover, functional annotation and KEGG pathway analysis were carried out to discover metabolic pathways, which are putatively influenced by the identified miRNAs and their target genes. Finally, the identified miRNAs and their targets were used for qPCR validation.

MATERIALS AND METHODS

Data Collection and Preparation

A total of 16,441 ESTs of *T. kok-saghyz* and 41,296 ESTs of *T. officinale* were obtained from National Center for Biotechnology

[DOI: 10.52547/jast.24.6.1457]

Information (NCBI) (http://www.ncbi.nlm.nih.gov) and used in this study. Total RNA-seq (paired-end short reads) data of both species were downloaded from the NCBI-SRA (Sequence Read Archive). Due to the effect of input data quality on the transcriptome assembly, the initial quality of raw reads was controlled for all samples using FastQC software (Version 0.11.9) (Mehta *et al*., 2016). Then, low-quality bases and reads and also possible adapter contaminants were removed using trimmomatic software by considering the minimum phred score (TRAILING) and minimum length (MINLEN) to 20 and 50, respectively (Bolger *et al*., 2014). After the trimming process, high quality reads (Clean reads) were used as input sequences for transcript assembling. Trinity software (v2.8.5), which is based on the strategy of de Bruijns graph, was used to assemble the raw reads (Haas *et al*., 2013). The quantity of *de novo* assembled transcriptomes was evaluated using Transrate software (v1.0.3) and some factors such as contig numbers, length of the largest contigs, N50, the average length of contigs were reported (Smith-Unna *et al*., 2016). In addition, Cap3 software was used to remove redundant contigs (Huang and Madan, 1999).

Identification of miRNAs

After *de nevo* assembly and removing the redundancy, the resultant transcripts and ESTs were used to identify conserved miRNAs. To predict miRNAs, C-mii software was used based on a homology search approach against known mature miRNAs (Numnark *et al*., 2012). This software detects the most similar transcripts to mature miRNA sequences using BLASTn by considering E-value ≤10 and mismatch ≤ 4 (Mehta *et al*., 2016). Mature miRNAs were used as queries and generated transcripts as subjects. The length range of potential miRNAs was determined to be between 19 and 24 nucleotides. Subsequently, the above transcripts were used in a BLSATx search against the NCBI database with an E-value cut-off of 10-5 to eliminate protein-coding sequences. To select candidate miRNAs, subsequent principles were used: locating mature miRNA in the arm of the hairpin with less than six mismatches, no miRNA-miRNA* duplex loops, and no more than 2 consecutive mismatches. Also, the predicted secondary structure must contain high negative Minimal Free Energy (MFE) and high Minimal Free Energy Index (MFEI) values with 30–70% A+U content (Singh *et al*., 2016; Mehta *et al*., 2016).

Target Prediction, Functional Annotation and Pathway Analysis

Prediction of target genes was performed using homology search between predicted conserved miRNAs and assembled transcriptome using C-mii software with the following criteria (Singh *et al*., 2016):

1) Mismatch more than four and especially at position 10th and 11th of the complementary site were not allowed, while only one mismatch at the position of 1-9 could exist,

2) For stable structure between miRNAs and their target, MFE value should be a negative,

3) No more than two consecutive mismatches were allowed.

Functional annotation of miRNAs target genes was carried out using BLASTx search against Uniprot-TrEMBL in C-mii software and NCBI Non-redundant protein database (Nr) with an E- value cut-off of 10-5 in Blast2GO software. By mapping sequences resulting from Blast2GO analysis to the KEGG database, pathway annotations of targets were obtained. (Kanehisa and Goto, 2000).

Gene Ontology Enrichment and Network Analysis

For further assessment of the function of the target transcripts in comparison with the

[DOI: 10.52547/jast.24.6.1457]

reference transcriptome, the gene ontology enrichment analysis was performed using Blast2GO and WEGO tools. Blast2GO software was used for obtaining the ontology terms of the reference transcriptome and miRNA targets (Conesa *et al*., 2005). For this purpose, total miRNA targets of both *T. kok- saghyz* and *T. officinale* merged together and enriched against the merged reference transcriptome of both species. WEGO (Web Gene Ontology Annotation Plot) was used for visualization of enriched GO terms ($P< 0.05$) (Ye *et al*., 2006). For network analysis, the top-hit blast results of the *Taraxacum* transcriptome against TAIR10 were used as GeneMANIA input (Warde-Farley *et al*., 2010). The resulting network was used in Cytoscape software (V 3.8.2) for visualization of gene-gene and miRNA-gene interactions (Su *et al*., 2014).

Total RNA Isolation and Real-Time PCR Analysis

T. kok-saghyz seeds were sterilized in ethanol 70% for 1 minute and Sodium hypochlorite (10 %) for 5 minutes and then were planted and grown in a greenhouse for 5 months. In the following, total RNA was isolated from the leaves, roots, stems, and flowers of the plant using RiboEx Total RNA reagent (GeneAll Biotechnology Co., Ltd., Songpa-gu, South Korea) according to the manufacturer's instruction and then exposed to DNaseI, RNase-Free DNase Set (Fermentase, Waltham, Massachusetts, USA). The quality and quantity of RNA samples were determined using agarose gel electrophoresis and Nanodrop ND-1000, respectively. Stem-loop RT and gene-specific real-time PCR primers were designed for five miRNAs (miR5021, miR838, miR1533, miR5015b, and miR1436) (Chen *et al*., 2005). (Supplementary Table S1). The miRNA stem-loop reverse transcription experiments were performed using the Superscript III First-Strand Synthesis System (Invitrogen) (Varkonyi-Gasic *et al*., 2007). qRT-PCR was performed on a Rotor-Gene 5- plex instrument based on the manufacturer's instructions. Briefly, 2.0 µL of diluted cDNA template was added to 3 µL of the Eva Green qPCR Master Mix 2x, 0.4 µL each primer and ddH2O to a final volume of 15 µL. The qPCR was amplified for 10 minutes at 95°C, followed by 45 cycles of 15 seconds at 95°C, 20 seconds at the specific annealing temperature for each primer, and 20 seconds at 72°C. *T. kok-saghyz* actin gene was used as an internal control. All the reactions were repeated at least three biological replications each by two technical replicates for statistical analysis, and the relative expression of genes was calculated using the REST software (Pfaffl *et al*., 2002).

RESULTS

Data Preparation, miRNAs Identification and Target Prediction

After integrating all RNA-seq database and eliminating low-quality reads, 178, 847,142 and 326, 152, 343 high-quality reads were obtained from *T. kok-saghyz* and *T. officinale*, respectively. Trinity software (v 2.8.5) produced 217,207 transcripts with an N50 of 1,782 bp in a range of 179 to 17,372 bp in *T. kok-saghyz* and 105,385 transcripts with an N50 of 1,506 bp in a range of 201 to 16,635 bp in *T. officinale* (Table 1). After removing redundant sequences using the Cap3 software, the number of contigs was reduced to 20,595 and 13,677 in *T. kok-saghyz* and *T. officinale*, respectively. ESTs and contigs were uploaded in C-mii software and identification of miRNAs was performed based on a homology search between miRNAs and the contigs and ESTs. After performing BLASTx and removing the protein-encoding sequences, the secondary structures of transcripts containing possible miRNAs were investigated by UNAFold software in C-mii (Xie *et al*., 2010;

[DOI: 10.52547/jast.24.6.1457]

Table 1. Summary statistics on de nevo assembly of *T. kok-saghyz* and *T. officinale.*

Length range (bp)	*Taraxacum kok-saghyz*	*Taraxacum officinal*
n_under_200	24	0
n_over_1k	85030	40382
n_over_10k	62	12
Total Number	217207	105385
N 50 (bp)	1782	1506
Mean Length (bp)	1090.22	1014.72
Smallest Length (bp)	179	201
Largest Lengh (bp)	17372	16635
N with orf	74610	41245

Hammond, 2015). Finally, conserved miRNAs in the *Taraxacum* genus were identified based on the percentage of AU and MFEI. Using RNA-seq and EST data, 574 miRNAs from 225 families in *T. kok-saghyz* and 396 miRNAs from 174 families in *T. officinale* were predicted by the homology search. In a similar study, which aimed to identify conserved miRNAs in *Stevia rebaudiana*, high efficiency was shown by using RNA-seq data and their *de novo* assembly to identify conserved miRNAs (Mehta *et al*., 2016). Our study showed that miR5021 was the most frequent in both species.

To detect miRNA target genes, we used C-mii software based on the homology search between mature identified miRNAs against the transcriptome and then were annotated against the Uniprot-TrEMBL database with an E-value cut-off of 10-5. According to the stated criteria, 192 putative targets for 77 miRNAs from 42 families in *T. kok-saghyz* and 174 putative targets for 50 miRNAs from 30 families in *T. officinale* were identified. The MFE of miRNA target secondary structures were ranged from −10.1 to −43.4 kcal mol^{-1} in *T. kok- saghyz* and -11 to -39.2 in *T. officinale*. For further analysis, functional annotation of miRNA target genes was performed using the BLASTX search against the non-redundant protein database of NCBI with an E-value cut-off of 10-5. According to the results, 77 and 83% of target genes, respectively in *T. kok-saghyz* and *T. officinale*, had significant homologs.

Gene Ontology Enrichment and Network Analysis

The purpose of the GO enrichment analysis is to find which GO terms are over-represented or under-represented. Based on GO analysis, 60 and 63% target genes, respectively for both *T. kok-saghyz* and *T. officinale*, were identified that had at least one significant phrase belonging to one of the three groups of biological processes, molecular functions, and cellular components. Gene ontology enrichment analysis showed that, in terms of molecular function, binding groups (51.5%), catalytic activity (36%), and transcription regulator activity (4.9%) were the most abundant and enriched in the target genes. This was while in the field of biological process, cellular process (54.9%), metabolic process (48.3%), and response to the stimulus (40.2%) were the most abundant and enriched in the targets. In addition, cell (57.8%), organelle (53.1%), and membrane (32.1%) components were the most abundant and enriched in the field of cellular components (Figure 1). Network analysis was performed based on the co-expression network of genes in *Arabidopsis thaliana* using GeneMANIA app in Cytoscape. As shown in Figure 2, contig17009 annotated as HMGCR (3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase), contig8094 annotated as DXS (1- Deoxy-D-Xylulose-5-phosphate Synthase), and contig2167 annotated as

[DOI: 10.52547/jast.24.6.1457]

RXN (2-C-methyl-D-erythritol 4-phosphate cytidylyl transferase) are regulated by miR5021. Contig17009 is also identified as a target for miR838. Furthermore, contig2103 annotated as PMVK (phosphomevalonate kinase) and contig2074 annotated as GPPSYN (Geranyl diPhosphate Synthase) are affected by miR5658 and contig489 annotated as ISPH2 (Isopentenyl-diPhosphate: NAD(P)+oxidoreductase) is

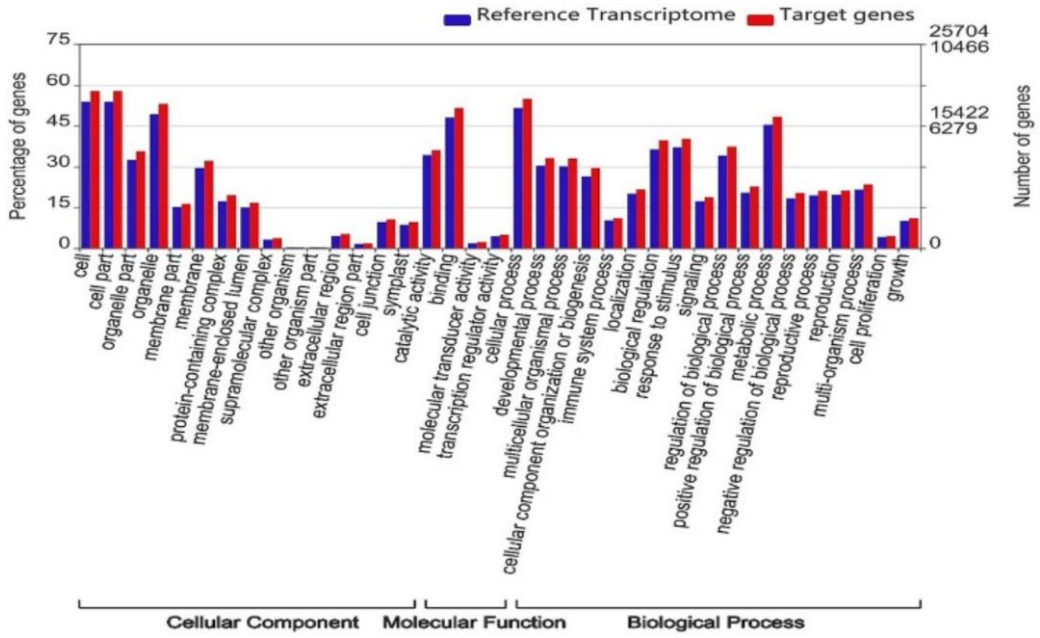

Figure 1. Gene ontology enrichment of the miRNA targets compared to the reference transcriptome of *Taraxacum* genus. Statistically significant (P< 0.05) enriched GO categories mainly under the cellular component, molecular function, and biological process categories were visualized by the WEGO tool.

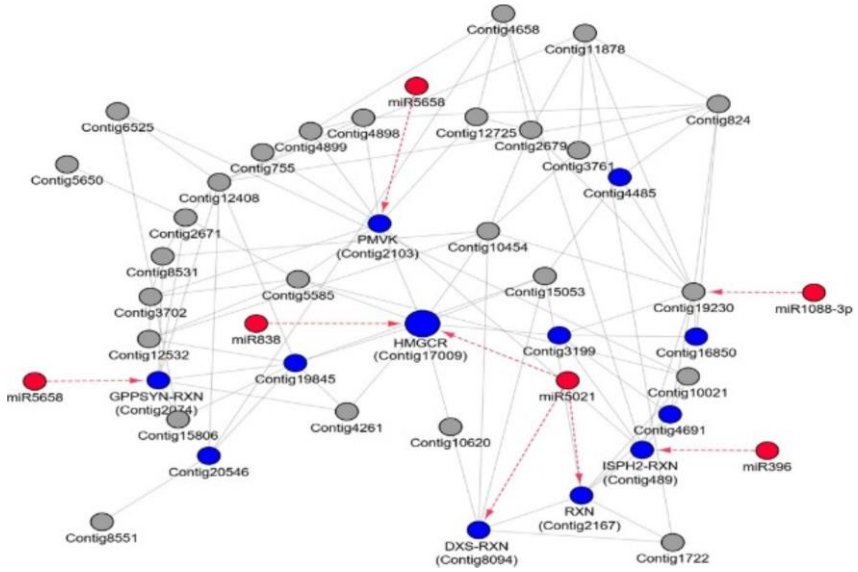

Figure 2. MicroRNA interaction network of terpene biosynthesis genes in *Taraxacum* genus. Blue nodes represent terpene biosynthesis genes, while grey nodes represent other genes. Red nodes represent the identified microRNAs in this study. Solid lines between gene nodes represent co-expression interaction between genes and dashed lines represent miRNA-gene interactions.

[DOI: 10.52547/jast.24.6.1457]

putatively under the influence of miR396.

Classification Based on KEGG Pathway Analysis Results

For a more detailed study of the biologically active pathways in *T. kok-saghyz* and *T. officinale*, the target genes were annotated with the Blast2GO against the KEGG database. One hundred and fifty-eight targets were assigned to 50 KEGG pathways using Blast2GO. According to the KEGG results, purine metabolism and thiamine metabolism were the most represented category.

qRT-PCR Confirmation of Predicted miRNAs

In this research, qRT-PCR was used to confirm the expression of miR5021 and miR838 (encoding by GO672159.1 and contig13182 respectively) and their target gene named HMGCR (contig17009, Accession Number: HQ857601.1). Additionally, miR5015b (encoding by Contig10984) and its target gene named invertase (contig 5863) was evaluated. The enzyme HMGCR is involved in the conversion of 3-Hydroxy-3-methylglutaryl-CoA to Mevalonate and it is a key enzyme in the MVA pathway (Figure 3). By considering leaf tissue as control, our results showed that miR5021 had the highest relative expression among all the investigated miRNAs in the root (Figure 4). MiR5021 and miR838 had higher relative expression in all tissues in comparison with the controls. MiR5021 and miR838, which both target *HMGCR* gene, had higher relative expression than their target genes in all tissues. The greatest rate of miR1533 transcripts was observed in the flower, while the stem had the lowest expression. According to the KEGG results, miR1533 potentially targets the *4-phosphate*

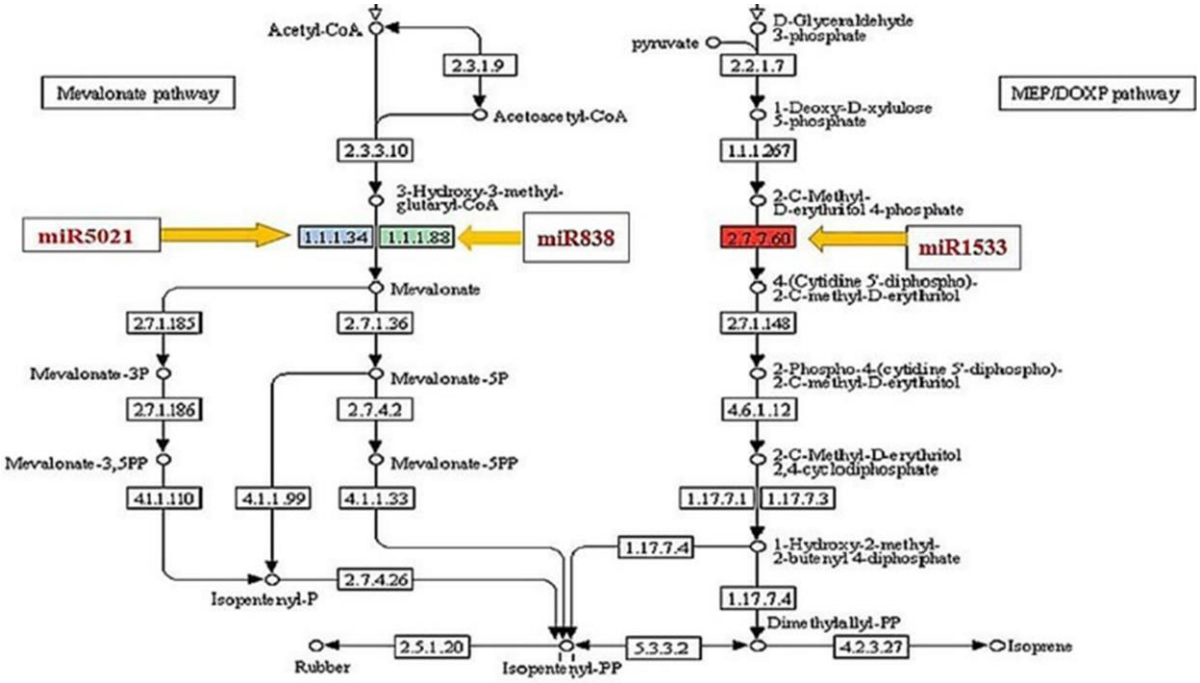

Figure 3. Schematic representation of terpene biosynthesis pathway in *T. kok-saghyz* showing that miR5021, miR838 and miR1533 interact with the *HMG-CoA reductase* and *4-phosphate cytidylyltransferase* genes, respectively. Top 20 KEGG pathways assigned for the targets predicted miRNA in *Taraxacum* genus.

[DOI: 10.52547/jast.24.6.1457]

cytidylyltransferase gene, respectively (Figure 3). The *4-phosphate cytidylyltransferase* gene catalyzes the formation of 4- diphosphocytidyl-2-C-Methyl-D-Erythritol from CTP and 2-C-methyl-D-erythritol 4-Phosphate (MEP) and is the plastid non-mevalonate pathway gene for isoprenoid biosynthesis. The maximum transcriptional activity of miR1533 and miR1436 occurred in the flower, while the highest transcripts levels of miR5015b were observed in the stem. MiR5015b had higher relative expression in root and stem tissues than the control, while miR1436 had higher relative expression in all tissues (Figure. 4). We have found that invertase and cellulose synthase are potential targets of miR5015b and miR1436, respectively. Invertase is a key enzyme in sucrose degradation. psRNATarget software was used to identify the effect of miRNAs related to terpenoids, starch and sucrose on their target genes (Dai and Zhao, 2011), and the results showed that all miRNAs target them by cleaving their target transcripts. A summary of miRNAs characterizations is available in Table 2 and the secondary structures of identified miRNAs are shown in Figure 5.

DISCUSSION

In recent years, with the emergence of NGS-based technologies, production of omics data in plants has grown dramatically (Singh *et al*., 2015), which allows identification of genes and factors involved in biological networks, including miRNAs, without the need for a reference genome sequence (Duan *et al*., 2012; Sabzehzari and Naghavi, 2019). So far, 19 miRNAs belonging to 7 families have been identified using EST data in *Taraxacum* (Srivastava *et al*., 2017), but the miRNAs of this plant remain largely unknown. Overall, in this study, 970 miRNAs from 399 families were identified in the *Taraxacum* using EST and RNA-seq data. Uracil has been detected to be the first nucleotide in the position of the 5′ end of the mature miRNAs as previously reported (Frazier *et al*., 2010). A wide range of miRNA precursor lengths from 35 to 2404 was found in *T. kok-saghyz* and 34 to 594 in *T. officinale*. The mean length of predicted miRNA precursors was 181.58 in *T. kok-saghyz* and 100.16 in *T. officinale* that is well coordinated with the result of other studies in plants (Singh and Sharma, 2014). The higher AU percentage ensures the stability and accuracy of the secondary structure of the predicted miRNA precursors. The precursor structure of miRNAs with a higher percentage of AU is richer in hydrogen bonds than the other forms of RNA. Hydrogen bonds often stabilize the hairpin structure of the miRNA series (Zhang *et al*., 2006a; Zhang *et al*., 2006b). In several studies, the MFEI parameter was the most appropriate criterion reported for miRNA identification (Mandhan *et al*., 2012). The MFEI calculated the secondary structure of the miRNAs ranging from -0.6 to -3.03 in *T. kok-saghyz* and -0.6 to -5.5 in *T. officinale*. The average MFEI calculated was -0.88 in *T. kok-saghyz* and -0.82 in *T. officinale*.

According to the results, it can be said that identified miRNAs in the present study are similar to the previous studies on miRNAs (Chen *et al*., 2005; Jung *et al*., 2013; Baldrich *et al*., 2018), play an important role in most biological processes and metabolism such as differentiation, growth and development, the transition from vegetative to reproductive stage, signaling and response to the stimulus. Our study showed that miR5021 was the most abundant in both species and potentially targets heat shock 70 protein, potassium channel NKT1, putative RING zinc finger protein, calcium-dependent calmodulin-independent protein kinase (CDPK), and Histone acetyltransferase. Highly abundant miRNAs in this study are involved in protein folding, regulation of transcription, and response to water deprivation and salt stress. While miR5021 plays an important role in response to

[DOI: 10.52547/jast.24.6.1457]

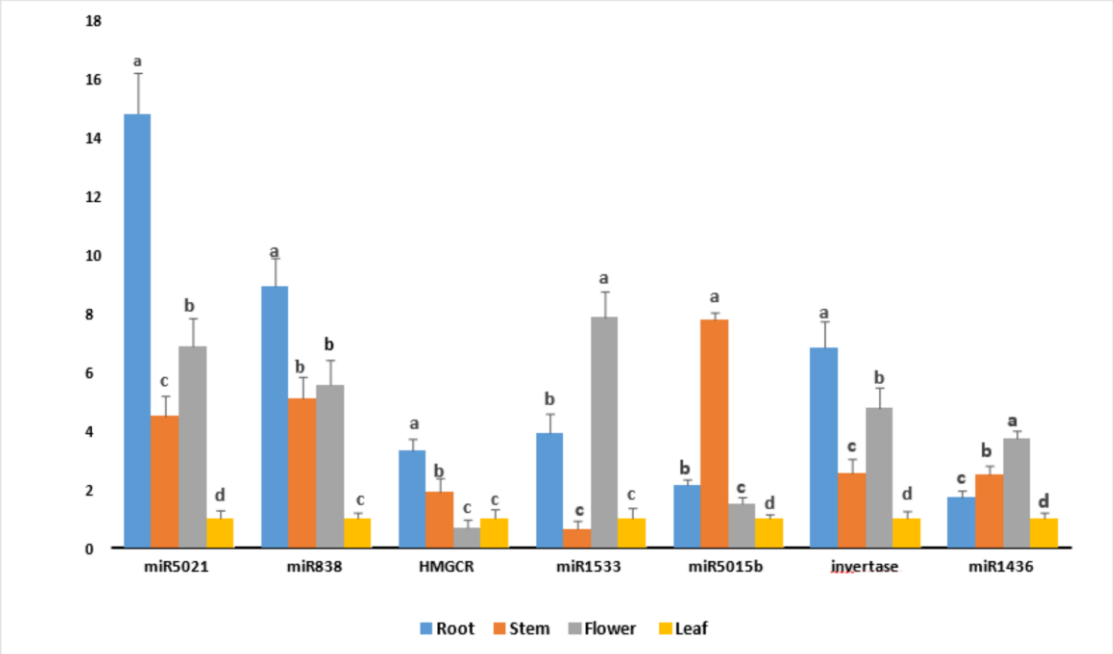

Figure 4. The relative expression level of predicted miRNAs and target genes in root, leaf, stem, and flower of *T. kok-saghyz*. Error bars represent SE (n= 3). Means followed by the same letter are not significantly different according to Duncan at 0.05 probability level.

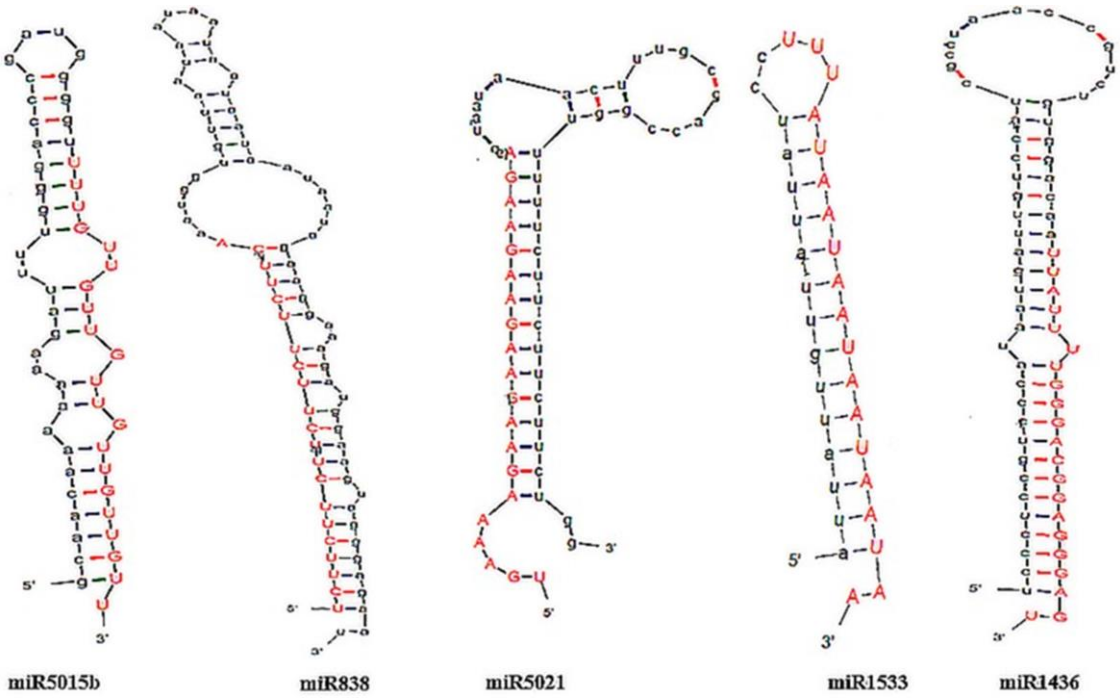

Figure 5. Secondary structure of predicted precursor miRNAs putatively involved in terpene, starch, and sucrose biosynthesis pathway.Red bases represent the miRNA sequences.

[DOI: 10.52547/jast.24.6.1457]

oxidative stress, regulation of transcription, transport, growth and development in *Catharanthus roseus* (Pani and Mahapatra, 2013). The KEGG pathway analysis results showed that the miR5021, miR838 and miR1533 are related to the terpene biosynthesis pathway. The role of miR5021 and miR838 has also been proven in the synthesis of terpene in other plants (Pani and Mahapatra, 2013; Fan *et al*., 2015; Singh *et al*., 2016; Najafabadi and Naghavi, 2018). Terpenes are produced from the DOXP/MEP and MVA pathways. According to the KEGG results, miR5021, miR838 and miR1533 potentially target the *HMGCR* and *4-phosphate cytidylyltransferase* genes, respectively (Figure 3). MiR5015b and miR1436 were found to be related to the pathway of starch and sucrose biosynthesis. MiR1436 was related to the phenylpropanoid pathway in *T. officinale* and targeted the *gentiobiase* and *lactoperoxidase* genes. The *gentiobiase* and *lactoperoxidase* genes are involved in the synthesis of coumarin and lignin, respectively. qRT-PCR analysis validated the expression level differences of the identified *T. kok-saghyz* miRNAs and the target genes in root, stem, flower and leaf tissue. The lower relative expression level of the miR5015b was consistent with the higher expression level of the *invertase* gene in root and flower. Also, the increase of miR5015b expression may lead to decreased expression of the invertase in stem. The important role of invertase in plant defense by activating plant immune responses against pathogens is well documented (Tauzin and Giardina, 2014). In addition, the results showed that the relative expression of miR5021 and miR838 was higher than that of the *HMGCR* gene in all tissue. The differences in the expression pattern of miR5021, miR838 and *HMGCR* gene suggests that miR5021 and miR838 play an important regulatory role in the terpenoid pathway.

CONCLUSIONS

MiRNAs act as one of the most important parts of regulatory networks in the plant and play a significant role by controlling the genes involved in the biosynthesis pathways. In this study, miRNAs of the *Taraxacum*

Table 2. Characterization of identified miRNAs involved in terpene, starch, and sucrose biosynthesis pathway.

miRNAs	Sequence name	Strand	Precursor	MFE (kcal mol^{-1})	MFEI (kcal mol^{-1})	mismatches GC%	Predicted	miRNA sequencelength
miR5021	GO672159.1	Minus	304	-72.3	-0.639	2	37.04	5':UGAAAAGAAGAAGAAGAAGA:3'
miR1436	Contig16591	Minus	250	-62.4	-0.77	3	32.27	5':UUAUUUUGGGACGGAGGGAGU:3'
miR5015b	Contig10984	Plus	796	-147.2	-0.64	4	28.85	5':UUUGUUGUUGUUGUUGUUU:3'
miR1533	Contig20226	Plus	34	-9.1	-3.03	4	8.57	5':UUUAUAAUAAUAAUAAUAA:3'
miR838	Contig13182	Plus	284	-86.4	-0.62	2	48.42	5':UUUUCUUCUUCUUCUUCUUUA:3'

[DOI: 10.52547/jast.24.6.1457]

Supplementary Table S1. List of primers used in this work (Karimi *et al*., 2020).

Gene	Seq
>ath-miR5021	UGAGAAGAAGAAGAAGAAAA
Forward Primer	CCATTGTTAATCGAGTGAGAAGAAGAAGA
Stem Loop	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACcagagccaACTTTTCT
>ath-miR838	UUUUCUUCUACUUCUUGCACA
Forward Primer	GACTGACAGTACGTTGTTCTATCTACTTCU
Stem Loop	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACcagagccaACTGTGCA
>ath-miR1533	AUAAUAAAAAUAAUAAUGA
Forward Primer	GACAGACTGACATGTGATAATAAGAATATAA
Stem Loop	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACcagagccaACTCATTA
>ath-miR5015b	UUCUGUUGUUGUUGGUGUUAU
Forward Primer	GATGGGGTTTTCTGTTGTTGTTGG
Stem Loop	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACcagagccaACATAACA
>hvu-miR1436	ACAUUAUGGGACGGAGGGAGU
Forward Primer	GATGGACAATACATTATGGGACGGA
Stem Loop	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACcagagccaACACTCCC
Universal Reverse Primer	GTGCAGGGTCCGAGGT
HMGCR	
Forward Primer	CCGTTTTCAACAAATCAAGCCGAT
Reverse Primer	ACCATGTTCATCCCCATTGCATC
Invertase	
Forward Primer	CGGACTCACGTTCCCAATAA
Reverse Primer	CACTGCGGTCAACAAGTTTAAT

genus were identified by the homology search approach. MiR1438 was involved in the biosynthesis of coumarin and lignin in the *T. officnale*. An integrating analysis of homology search and network analysis showed that miR5021 and miR838 were putatively involved in HMGCR regulation. Experimental validation of miR5021 and miR838 indicated a higher relative expression in all tissues in comparison with their target gene in *T. kok-saghyz*, indicating that it is actively regulated by miR5021 and miR838. There was a positive correlation between the quantity of rubber and the relative expression of genes involved in rubber synthesis, including the *HMGCR* gene in *T. kok-saghyz* (Panara *et al*., 2018). Therefore, miR5021 and miR838 play a key role in regulating rubber production by affecting the *HMGCR* gene. This finding will provide a new path to the breeding program of rubber production in *T. kok-saghyz*. However, due to the negative impact

[DOI: 10.52547/jast.24.6.1457]

of miR5021 and miR838 on its target gene, dsRNA technology in genetic engineering can be used to shut down miR5021 and miR838 to improve the quantity and quality of rubber in *T. kok-saghyz.*

Abbreviations list:

allylic-PP: allylic Pyrophosphate
Bp: Base pair(s)
BLAST: Basic Local Alignment Search Tool
CPTs: Cis-Prenyltransferases
DOXP/MEP: 2-C-1-Deoxy-D-Xylulose-5-Phosphate/Methyl-D-Erythritol-4-Phosphate
DXS: 1-Deoxy-D-Xylulose-5-phosphate Synthase
E-value: Expectation value
EST: Expressed Sequence Tag
GO: Gene Ontology
GPPSYN: Geranyl diPhosphate Synthase
HMGCR: 3-Hydroxy-3-Methylglutaryl-CoA Reductase
IPP: Isopentenyl Pyrophosphate
ISPH2: Isopentenyl-diPhosphate: NAD(P)+oxidoreductase
KEGG: Kyoto Encyclopedia of Genes and Genome
MVA: Mevalonate pathway
MFE: Minimal Free Energy
MFEI: Minimal Free Energy Index
NR: Natural Rubber
PIEs: Phenolic Inositol Esters
PMVK: Phosphomevalonate Kinase
RXN: 2-C-methyl-D-erythritol 4-phosphate cytidylyl transferase
REF: Rubber Elongation Factor
SRPPs: Small Rubber Particle Protein
TritAc: Triterpene Acetates
TF: Transcription Factor
WEGO: Web Gene Ontology Annotation Plot.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the University of Tehran and Iran National Science Foundation for the financial support of this work.

REFERENCES

1. Baldrich, P., Beric, A. and Meyers, B.C. 2018. Despacito: The Slow Evolutionary Changes in Plant MicroRNAs. *Curr. Opin. Plant Biol*., **42**: 16-22.
2. Bolger, A.M., Lohse, M. and Usadel, B. 2014. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics*. **30**: 2114-2120.
3. Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., Barbisin, M., Xu, N. L., Mahuvakar, V. R. and Andersen, M. R. 2005. Real-Time Quantification of MicroRNAs by Stem–Loop RT–PCR. *Nucl. Acids Res*., **33**: e179-e179.
4. Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M. and Robles, M. 2005. Blast2GO: A Universal Tool for Annotation, Visualization and Analysis in Functional Genomics Research. *Bioinformatics*, **21**: 3674-3676.
5. Dai, X. and Zhao, P. X. 2011. psRNATarget: A Plant Small RNA Target Analysis Server. *Nucl. Acids Res.,* **39:** W155-W159.
6. Duan, J., Xia, C., Zhao, G., Jia, J. and Kong, X. 2012. Optimizing de Novo Common Wheat Transcriptome Assembly Using Short-Read RNA-Seq Data. *BMC Genom.,* **13**: 392.
7. Fan, R., Li, Y., Li, C. and Zhang, Y. 2015. Differential MicroRNA Analysis of Glandular Trichomes and Young Leaves in *Xanthium strumarium* L. Reveals Their Putative Roles in Regulating Terpenoid Biosynthesis. *PLoS One*, **10**: e0139002.
8. Frazier, T.P., Xie, F., Freistaedter, A., Burklew, C. E. and Zhang, B. 2010. Identification and Characterization of MicroRNAs and Their Target Genes in Tobacco (*Nicotiana tabacum*). *Planta*, **232**: 1289-1308.
9. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B. and Lieber, M. 2013. De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Nat. Protoc*., **8:**

[DOI: 10.52547/jast.24.6.1457]

1494-1512.
10. Hammond, S. M. 2015. An Overview of MicroRNAs. *Adv. Drug Deliv. Rev.,* **87:** 3-14.
11. Hayashi, Y. 2009. Production of Natural Rubber from Para Rubber Tree. *Plant Biotechnol. J.*, **26**: 67-70.
12. Huang, X. and Madan, A., 1999. CAP3: A DNA Sequence Assembly Program. *Genome Res*., **9:** 868-877.
13. Huber, M., Triebwasser-Freese, D., Reichelt, M., Heiling, S., Paetz, C., Chandran, J. N., Bartram, S., Schneider, B., Gershenzon, J. and Erb, M. 2015. Identification, Quantification, Spatiotemporal Distribution and Genetic Variation of Major Latex Secondary Metabolites in the Common Dandelion (*Taraxacum officinale* agg.). *Phytochemistry*, **115**: 89-98.
14. Jung, H. J., Park, S. J. and Kang, H. 2013. Regulation of RNA Metabolism in Plant Development and Stress Responses. *J. Plant Biol.,* **56**: 123-129.
15. Kanehisa, M. and Goto, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res*., 28: 27-30.
16. Karimi, A. A., Naghavi, M. R., Peyghambari, S. A. and Rasoulnia, A. 2021. Inulin Content and Expression of Related Genes in Different Tissues and Cell Suspension Culture of *Taraxacum kok-saghyz. In Vitro Cell. Dev. Biol. Plant*, **57**: 1009–1017.
17. Li, L., Xu, J., Yang, D., Tan, X. and Wang, H. 2010. Computational Approaches for MicroRNA Studies: A Review. *Mamm. Genome*, **21:** 1-12.
18. Luo, Z., Iaffaldano, B.J., Zhuang, X., Fresnedo-Ramírez, J. and Cornish, K. 2017. Analysis of the First *Taraxacum kok-saghyz* Transcriptome Reveals Potential Rubber Yield Related SNPs. *Sci. Rep.,* **7:** 1-13.
19. Mandhan, V., Kaur, J. and Singh, K. 2012. smRNAome Profiling to Identify Conserved and Novel MicroRNAs in *Stevia rebaudiana* Bertoni. *BMC Plant Biol.,* **12**: 197.Mathiyalagan, R., Subramaniyam, S., Natarajan, S., Kim, Y. J., Sun, M. S., Kim, S. Y., Kim, Y. -J. and Yang, D. C. 2013. *In Silico* Profiling of MicroRNAs in Korean Ginseng (*Panax ginseng* Meyer). *J. Ginseng Res.*, **37**: 227.
20. Mehta, A., Gupta, H., Rawal, R., Mankad, A., Tiwari, T., Patel, M. and Ghosh, A. 2016. *In Silico* MicroRNA Identification from *Stevia rebaudiana* transcriptome Assembly. *Eur. J. Med. Plants*, **15(2):** 1-14.
21. Mooibroek, H. and Cornish, K. 2000. Alternative Sources of Natural Rubber. *Appl. Microbiol. Biot*., **53**: 355-365.
22. Najafabadi, A. S. and Naghavi, M. R. 2018. Mining *Ferula gummosa* Transcriptome to identify miRNAs Involved in the Regulation and Biosynthesis of Terpenes. *Gene*, **645**: 41-47.
23. Nowicki, M., Zhao, Y., Boggess, S. L., Fluess, H., Payá-Milans, M., Staton, M. E., Houston, L. C., Hadziabdic, D. and Trigiano, R. N. 2019. *Taraxacum kok-saghyz* (Rubber Dandelion) Genomic Microsatellite Loci Reveal Modest Genetic Diversity and Cross-Amplify Broadly to Related Species. *Sci. Rep.*, **9**: 1-17.
24. Numnark, S., Mhuantong, W., Ingsriswang, S. and Wichadakul, D. 2012. C-mii: A Tool for Plant miRNA and Target Identification, *BMC Genom.,* Vol. **13**, Article Number: S16.
25. Panara, F., Lopez, L., Daddiego, L., Fantini, E., Facella, P. and Perrotta, G. 2018. Comparative Transcriptomics between High and Low Rubber Producing *Taraxacum kok-saghyz* R. Plants. *BMC Genom.,* **19:** 1-14.
26. Pani, A. and Mahapatra, R.K., 2013. Computational Identification of MicroRNAs and Their Targets in *Catharanthus roseus* Expressed Sequence Tags. *Genom. Data*, **1:** 2-6.
27. Patel, S. and Goyal, A. 2012. The Current Trends and Future Perspectives of Prebiotics Research: A Review. 3 *Biotech*, **2**: 115-125.
28. Pfaffl, M. W., Horgan, G. W. and Dempfle, L. 2002. Relative Expression Software Tool (REST©) for Group-Wise Comparison and Statistical Analysis of Relative Expression Results in Real-Time PCR. *Nucl. Acids Res.,* **30**: e36-e36.
29. Ramirez-Cadavid, D. A., Cornish, K. and Michel Jr, F. C. 2017. *Taraxacum Kok-saghyz* (TK): Compositional Analysis of a Feedstock for Natural Rubber and Other Bioproducts. *Ind. Crop Prod.,* **107**: 624-640.
30. Sabzehzari, M. and Naghavi, M. R. 2019. Phyto-miRNA: A Molecule with Beneficial Abilities for Plant Biotechnology. *Gene*, **683**: 28-34.

[DOI: 10.52547/jast.24.6.1457]

31. Salehi, M., Bahmankar, M., Naghavi, M. and Cornish, K. 2022. Rubber and Latex Extraction Processes for *Taraxacum kok-saghyz*. *Ind. Crop Prod.*, **178:** 114562.
32. Samad, A. F., Sajad, M., Nazaruddin, N., Fauzi, I.A., Murad, A., Zainal, Z. and Ismail, I. 2017. MicroRNA and Transcription Factor: Key Players in Plant Regulatory Network. *Front. Plant Sci.*, **8**: 565.
33. Singh, N. and Sharma, A. 2014. *In-Silico* Identification of miRNAs and Their Regulating Target Functions in *Ocimum basilicum*. *Gene*, **552**: 277-282.
34. Singh, N., Srivastava, S., Shasany, A. K. and Sharma, A. 2016. Identification of miRNAs and Their Targets Involved in the Secondary Metabolic Pathways of *Mentha spp*. Comput. Biol. Chem., **64:** 154-162.
35. Singh, V. K., Singh, A. K., Singh, S. and Singh, B. D. 2015. Next-Generation Sequencing (NGS) Tools and Impact in Plant Breeding. In: "*Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*". Springer, PP. 563-612.
36. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M. and Kelly, S. 2016. TransRate: Reference-Free Quality Assessment of de Novo Transcriptome Assemblies. *Genome Res.*, **26:** 1134-1144.
37. Srivastava, S., Singh, N., Srivastava, G. and Sharma, A. 2017. MiRNA Mediated Gene Regulatory Network Analysis of *Cichorium intybus* (Chicory). *Agri. Gene*, **3:** 37-45.
38. Su, G., Morris, J.H., Demchak, B. and Bader, G. D. 2014. Biological Network Exploration with Cytoscape 3. *Curr. Protoc. Bioinformatics*, **47:** 8.13. 1-8.13. 24.
39. Tauzin, A. S. and Giardina, T. 2014. Sucrose and Invertases, a Part of the Plant Defense Response to the Biotic Stresses. *Front. Plant Sci.*, **5(293):** 1-8.
40. Tian, T., Wang, J. and Zhou, X. 2015. A Review: MicroRNA Detection Methods. *Org. Biomol. Chem.*, **13**: 2226-2238.
41. Van Beilen, J. B. and Poirier, Y. 2007. Establishment of New Crops for the Production of Natural Rubber. *Trends Biotechnol.*, **25**: 522-529.
42. Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E. F. and Hellens, R. P. 2007. Protocol: A Highly Sensitive RT-PCR Method for Detection and Quantification of MicroRNAs. *Plant Methods*, **3**: 1-12.
43. Wang, L., Zheng, Y., Ding, S., Zhang, Q., Chen, Y. and Zhang, J. 2017. Molecular Cloning, Structure, Phylogeny and Expression Analysis of the *Invertase* Gene Family in Sugarcane. *BMC Plant Biol.*, **17**: 1-20.
44. Wang, M., Wang, Q. and Wang, B. 2012. Identification and Characterization of MicroRNAs in Asiatic Cotton (*Gossypium arboreum L.*). *PLoS One*, **7:** e33696.
45. Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F. and Lopes, C. T. 2010. The GeneMANIA Prediction Server: Biological Network Integration for Gene Prioritization and Predicting Gene Function. *Nucl. Acids Res.*, **38**: W214-W220.
46. Xie, F., Frazier, T. P. and Zhang, B. 2010. Identification and Characterization of MicroRNAs and Their Targets in the Bioenergy Plant Switchgrass (*Panicum virgatum*). *Planta*, **232**: 417-434.
47. Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R. and Bolund, L. 2006. WEGO: A Web Tool for Plotting GO Annotations. *Nucl. Acids Res.*, **34**: W293-W297.
48. Zhang, B., Pan, X., Cannon, C. H., Cobb, G. P. and Anderson, T. A. 2006a. Conservation and Divergence of Plant MicroRNA Genes. Plant J., **46**: 243-259.
49. Zhang, B., Pan, X., Cox, S., Cobb, G. and Anderson, T., 2006b. Evidence that MiRNAs Are Different from Other RNAs. *Cell Mol. Life Sci.*, **63:** 246-254.

[DOI: 10.52547/jast.24.6.1457]

شناسایی miRNAها و ژن‌های هدف آنها در گیاهان خانواده *Taraxacum*

ع. ا. کریمی، م. ر نقوی، س. ع. پیغمبری، ا. سبحانی و ع. رسول نیا

چکیده

میکروRNAها یک گروه از RNAهای غیر کدکننده پروتئینی درون زاد بوده که با تاثیر گذاری بر بیان ژن های هدف، نقش های حیاتی در تمام فرآیندهای متابولیک سلولی دارند. با وجود شناسایی miRNA ها در این خانواده، تا به امروز miRNA ها در گونه های *Taraxacum*، که یک گیاه صنعتی مهم است، تا حد زیادی ناشناخته مانده است. در مطالعه حاضر، با انجام رویکردهای محاسباتی، ۹۷۰ miRNA از ۳۹۹ خانواده در گونه های *Taraxacum* شناسایی شده است. miR5021 رایج ترین miRNA در گونه *Taraxacum* است. با توجه به نتایج KEGG، miR5021، miR838 و miR1533 به مسیر بیوسنتز ترپن مرتبط بودند، در حالی که miR5015b و miR1436 در مسیرهای بیوسنتز نشاسته و ساکارز نقش داشتند. آزمایش qPCR برای تایید سطوح بیان پنج miRNA و ۳-هیدروکسی-۳- متیل گلوتاریل کوآنزیم A ردوکتاز HMGCR () و اینورتاز به عنوان ژن های هدف انجام شد. نتایج نشان داد که بیشترین بیان نسبی miR1533 و miR1436 در گل رخ داد، در حالی که بالاترین سطح رونوشت miR5015b در ساقه مشاهده شد. علاوه بر این، سطح بیان نسبی بالاتر miR5021 و miR838 با سطح بیان پایین‌تر ژن HMGCR در تمام بافت‌ها مطابقت داشت، که نشان می‌دهد miR5021 و miR838 در تنظیم بیان ژن HMGCR نقش دارند. از آنجایی که مسیر موالونات منبع اصلی ایزوپنتنیل پیروفسفات است که در سنتز کائوچو استفاده می شود، پس miR5021 و miR838 با تنظیم بیان آنزیم HMGCR نقش مهمی در تولید کائوچو دارند. این یافته ها مطالعات چشم انداز آینده را در مورد مکانیسم های تنظیمی miRNA ها در *Taraxacum kok-saghyz* سرعت می بخشد.

[DOI: 10.52547/jast.24.6.1457]