Life Science Informatics Publications
Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences
Journal Home page http://www.rjlbpcs.com/
RJLBPCS
ISSN 2454-6348

Original Research Article **DOI - 10.26479/2017.0301.13**

IDENTIFICATION OF APRICOT MICRORNA FROM EST DATA AND THEIR CROSS-KINGDOM TARGETS IN HUMAN

Drushti H. Bhatt[1], Rakesh M. Rawal[2], Kaid Johar S. R[3], Himanshu A. Pandya[*1]

1 Department of Botany, Bioinformatics and Climate change Impacts Management, Gujarat University, Ahmedabad, GJ, India.

2 Department of Life Science, School of sciences, Gujarat University, Ahmedabad, GJ 380009, India.

3 Department of Zoology, School of sciences, Gujarat University, Ahmedabad, GJ 380009, India.

ABSTRACT: MicroRNAs (miRNAs) are small non-coding post-transcriptionally regulatory RNA, present in both plant and human. The food derived miRNAs have been proved to target genes in human and other animals and regulate their expression at genetic level. Based on this concept, we have computationally identified potential miRNAs of *Prunus armeniaca* (Apricot) from EST data. After performing different alignments and secondary structure identifications, the 6 (Six) potential miRNAs out of 17686 EST of apricot were identified and they were further subjected against human transcript to find out functional targets. The functional annotation and gene ontology of human target genes were carried out by bioinformatic tools and software. We have noticed that 2 miRNAs of apricot have higher complementarity with 27 human genes and these genes are involved in molecular functions (27%), biological processes (59%) and cellular components (14%).These finding expand the scope of understanding the functions of dietary miRNAs in human and give new insights to the researchers for experimental validation.

KEYWORDS: *Prunus armeniaca* (Apricot), MicroRNAs (miRNAs), Expressed Sequence Tags (ESTs)

***Corresponding Author: Prof. Dr. Himanshu A. Pandya** Ph.D.

Department of Botany, Bioinformatics and Climate change Impacts Management, Gujarat University, Ahmedabad, GJ, INDIA

* Email Address: hapandya24@yahoo.com

1. INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs that are around 20-25 nucleotides in length and can regulate several different genes through sequence-specific hybridization to the 3'- translated region (UTR) of messenger RNAs [1]. The miRNA gene is transcribed to form a primary miRNA (pri-miRNA) precursor molecule that goes through nuclear cleavage to generate precursor microRNA (pre-miRNA). The pre-miRNA is cleaved in the cytoplasm to create a miRNA duplex (miRNA:miRNA*), passenger strand designed with asterisk containing the mature miRNA. The duplex unwinds and the mature miRNA accumulates into RISC (RNA induced silencing Complex). The miRNA base-pairing with target mRNA direct gene silencing by mRNA cleavage or translation repression based on the complementarity between the miRNA and the mRNA target [2]. There is no need for prefect complementary of miRNAs with target mRNA for target recognition, so one miRNA can target multiple messenger RNAs. miRNAs play key role in several biological processes like immune modulation, metabolic control, neuronal development, cell cycle, muscle differentiation and stem cell differentiation. Almost all miRNAs are conserved across multiple species, showing the evolutionary importance of these molecules as modulators of essential biological pathways and processes [1,2]. Vegetables and fruits are valuable resource of healthy diet with a broadly recognized beneficial effect on human well-being. The number of miRNAs in corn, soybean seeds and rice has high sequence similarity to various animal genes. Some putative plant miRNAs could selectively package into microvesicles (MVs), which are small vesicles that are shed from almost all cell types under both normal and pathological conditions and delivered into recipient cells where the exogenous miRNAs can regulate target gene and recipient cell functions [3]. For the first time, *in-vitro* and *in-vivo* studies performed on food-derived plant miRNA, MIR168a of rice have shown that plant miRNA can pass through the mouse gastrointestinal (GI) track and enter the circulation and various organs especially the liver where it regulates mouse/human LDLRAP1 protein expression and physiological condition [3]. Second study of plant derived miRNAs was carried out on broccoli. The miR160 of broccoli down regulate genes that are significantly and frequently up regulated in non-small-cells lung cancer (NSCLC) [4]. These cross-kingdom regulation studies of miRNA and mRNA proved that these small riboregulators have potential to modulate the physiology of human genes. Therefore, we have design a study to computationally identify potential miRNAs from the *Prunus armeniaca* (apricot) and to find out their targets in human. *Prunus armeniaca* (Apricot) is trees classified under Rosaceae family and majorly of these plants grow in Korea, India, China, Iran, North Africa and Japan. Apricot is considered best sources of protein, oil, fiber, phenolic and cyanogenic compounds. It has been utilized in gastric inflammations, dermatitis and as a carminative agent. Pharmacologic studies have shown antioxidant and free radical scavenging properties, antitussive effects and antimicrobial

activity [5]. Considering above benefits, we have found that effect of miRNAs of apricot in human health has not been identified and therefore, we have predicted miRNAs computationally using EST approach and discovered their functional human target genes.

2. MATERIALS AND METHODS

The identification of miRNA of *Prunus armeniace* (Apricot) was performed by EST approach. The ESTs of apricot has been taken from NCBI dbEST [6]. There will be higher probability to obtain novel miRNAs of apricot if we take mature miRNAs from same family that is *Rosaceae* (*Molus domestica* and *Prunus persica*) as reference set for comparison with EST of apricot. The mature miRNAs of *Molus domestica* and *Prunus persica* has been retrieved from miRBASE database [7]. The duplication or repetitive sequences of reference set of mature miRNAs were removed and formed unique set for further evaluation. After fetching EST of apricot and reference miRNAs, we have performed local alignment by using BLASTn tool in LINUX platform. To improve search, expected value set to 1000 to increase number of potential hits; max target sequence was set to 6 (Six), number of mismatch was set to 4 (Four), word size was 7 (Seven) for initial matching between query and database and number of threads 3 (Three). After running BLASTn, we have selected those candidate sequences which have more than 18 (Eighteen) miRNA sequence length. These unique miRNA coordinates have been achieved after running BEDTOOLs. The only ESTs having these miRNA coordinates were selected and retrieved to compare with Protein Database (non-redundant) using BLASTx NCBI tool. As miRNAs are small non-coding RNAs, BLASTx was used to identify coding sequences from the ESTs, which we have further removed and only non-protein coding sequences left for next step. Then, we have manually observed non-protein coding sequences from the ESTs of apricot and made precursor miRNAs (pre-miRNAs) by taking 100nt from upstream and downstream from the location of predicted miRNAs of apricots. The secondary structure of pre-miRNAs was generated using the Zuker folding algorithm with web based computational software MFOLD, which was publicly available at http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form [8]. The only stem-loop structure of pre-miRNAs has been selected and used as novel miRNAs of apricot. The miRNA sequences present in stem part used as novel mature miRNAs of apricot and further taken as query to identify their mRNAs (target gene) in human by psRNATarget tool [9]. The detail study of these predicted targets were conducted by Gene Ontology Consortium and Human gene and protein database [10 and 11]. The sequence similarity between reported human miRNAs of particular target and predicted miRNAs of apricot has been performed by LocARNA tool to know the potency and feasibility of predicted miRNAs [12]. The reported human miRNAs of target genes were taken from Target Scan Human database [13] [Figure 1].

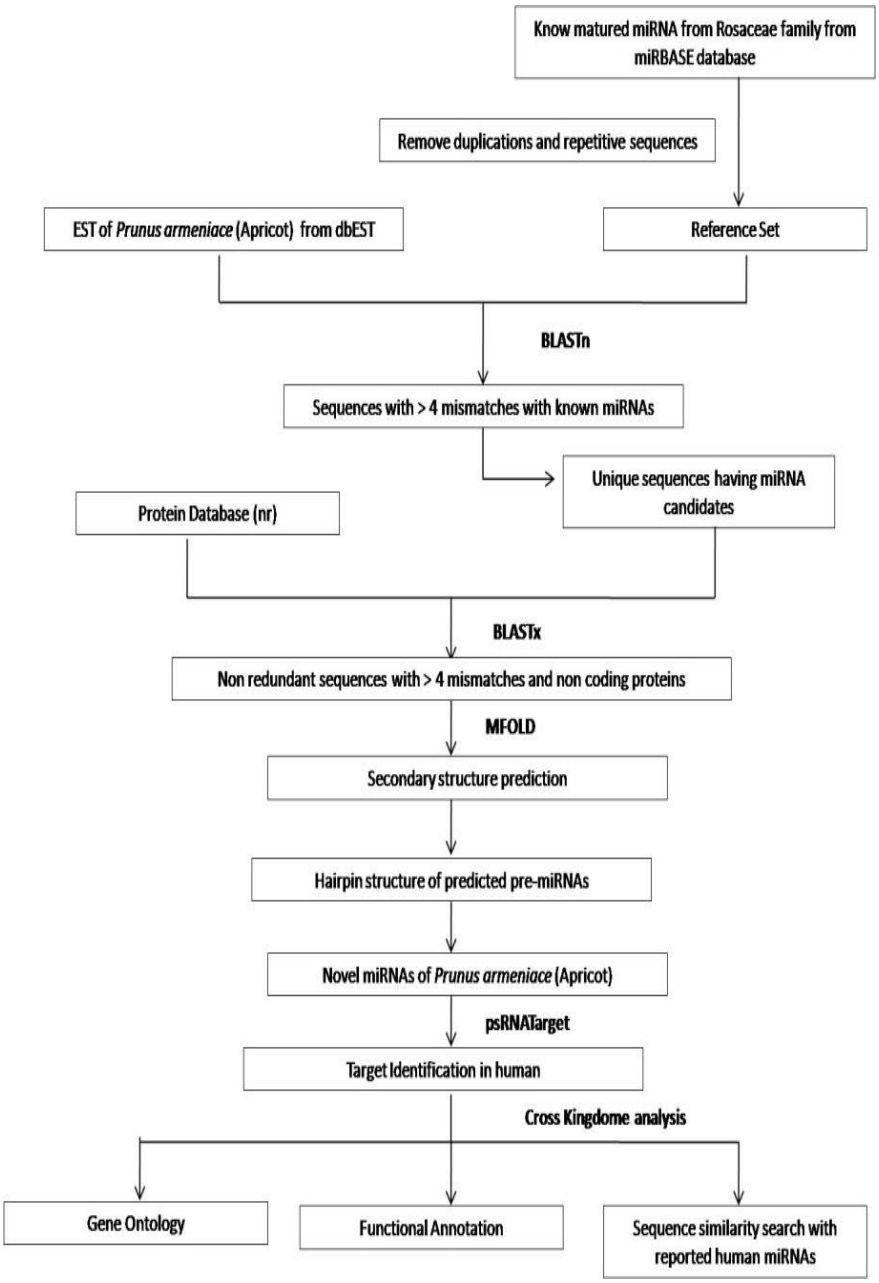

Figure 1. Workflow

3. RESULTS AND DISCUSSION

Retrieval of EST of *Prunus armeniace*(Apricot)

In this work, we have taken ESTs (Expressed Sequence Tags) of *Prunus armeniaca*(Apricot) to identify the potential miRNAs. ESTs are fragments of mRNA sequences derived through single sequencing reaction performed on randomly selected clones from cDNA libraries. They may be used to detect gene transcripts and are helpful in gene discovery and sequence determination. There are 17686 ESTs of apricot are available till date in NCBI dbEST [Table 1]. We have created database of these EST for local alignment with reference set.

Table 1: Statistics of work

No. of EST of *Prunus armeniaca* (Apricot)	17686
Mature miRNAs of *Rosaceae* (Reference Set)	421
miRNA homologues after BLASTn	128
Unique EST sequences after BLASTx	8
Predicted novel miRNAs based on secondary structure prediction	6
Functional miRNAs after predicting human targets	2
No. of human targets	27

Creation of Reference set

All mature miRNAs of *Rosaceae* family including *Molus domestica* and *Prunus persica* are previously sequenced and available in miRBASE database. Total 421 mature miRNAs were downloaded and removed repetitive sequences to avoid repeat search of redundant miRNAs [Table 1]. All these mature miRNAs defined as reference set for further procedure.

Identification of novel miRNA from EST of apricot

Conservation of mature miRNAs from species to species has made work easier to identify conserved miRNAs from the EST data. In first step, All ESTs were subjected to BLASTn against the reference set of miRNAs to identify miRNA homologues of apricot. After performing BLASTn, we have achieved 128 homologues for further investigation. These miRNAs homologues were blasted against protein database (nr) using BLASTx NCBI for identifying coding sequences of ESTs of apricot. At the end, we have found 8 (Eight) unique EST of apricot which has less than 4 mismatches with mature miRNAs and non-protein sequences [Table 2].

Table 2: Sequences of predicted miRNAs of Apricot and their MFE.

Predicted miRNAs of Apricot	miRNA sequences	Minimum Free Energy (Kcal/mol)
BQ134646	TTAGAGAGAGAGAGAGAGAG	-22.30
CB821048	TTTTTTTTTTAATTGAT	-24.79
CV047319	GTGCTATCCTTGCTGAGCTT	-21.20
CV047402	GAATAGTGTTTGCCATGG	-35.40
CV052881	GAATCATCAAGATTCAC	-20.80
DY653876	AAAAGGGAGGGGGAAAA	-29.20

After removing coding sequences, 8 EST were subjected to identify secondary structure prediction by MFOLD. Before that, we have manually taken 50 nt from upstream and 50 nt from downstream from the location of miRNAs found in ESTs of apricot. During procedure of secondary structure prediction, we have observed that length of predicted pre-miRNAs of apricot lied between 71-141nt. Out of them, 0-4 mismatches were found between predicted miRNAs and reference set.

Predicted pre-miRNAs have stem-loop structure without any breaks and loops within stem part and minimum free energy (MFE) of pre-miRNAs of apricot was higher than -20 Kcal/mol [Figure 6]. We have found 6 (six) novel candidate miRNAs, which are possessing all set criteria [Table 2]. In the process of secondary structure prediction, we have noticed that pre-miRNA like CV047402 and DY653876 required -35.40Kcal/mol and -29.20Kcal/mol MFE energy respectively [Table 2].

Identification of potential targets from EST derived miRNAs and their functional annotation

The target prediction was done based on perfect or near perfect complementarity with miRNAs and their mRNA targets. All 6 (Six) mature miRNAs of apricot was used for searching against mRNA sequences of human using web based server psRNATarget with default parameter. We have achieved 27 targets that hybridize with predicted miRNAs of apricot. During this procedure, we have noticed that only 2 (Two) miRNAs out of 6 (Six) were showing interaction with targets even if their pre-miRNAs were having required energy i. e. -22.30Kcal/mol and -21.2Kcal/mol respectively which are higher compared to CV047402 and DY653876 [Table 3]. The functional analysis and pathways analysis of 27 targets has been performed by Gene Ontology Consortium. The targets were involved in 19 molecular functions, 42 biological processes and 10 cellular components [Figure 2].

Table 3: List of predicted miRNAs of apricots with their human targets.

miRNAs of apricot	miRNA sequences	Target name	Target sequences	Target_start	Target_end
CV047319.1: 37-57	GUGCUAUCCUUGCU GAGCUU	SIRT5	AACUUCAGUAGGGAUGGCAC	1129	1148
		APLN	AAGUGCAGCAGGAAUAGCAC	1444	1463
Q134646.1:100-120	UUAGAGAGA GAGAGAGAGAG	CREB3L2	CUCUCCCUCUCUCUCUCUGU	1511	1530
		LENG8	UUCUCUCUCUCUCUUUCGAG	302	321
		LHFPL5	CACUCUCUCUCUCUCUCAAA	511	530
		TAOK3	CUCACUCUCUCUCUCUCUUA	101	120
		BEST3	CUCUUUCUUUUUUUCUUUAG	404	423
		TMEM135	CCCUCUCUCUCUCUCUCUGU	4169	4188
		PLRG1	CUCUUUUUUUUUCUUUUUAA	22	41
		YPEL5	CUUUCUUUCUUUUUUUUUAA	146	165
		HCFC1	CUCUUUCUCUCUGUUUUUAA	136	155
		PHF2	CUCUCUCUCUCUUUUUUUUA	1723	1742
		IFIT5	CUUUCUUUUUUUCUUUUUAA	181	200
		SPATA6L	CUCUCUGUCUCUCUCUCUCA	332	351
		CSNK1G3	CUUUUUUUUUUUUUCUCUAA	164	183
		ZNF507	UUUUUUUUCUUUCUUUCUAA	1626	1645
		DSTYK	CUUUCUCUUUCACUCUCUAG	11	30
		KDM5A	CUGUCUCUCUCUUUUUUUAA	3271	3290
		XYLT1	CUCUCUCUCUUCCUCUCUGG	4727	4746
		RBMS3	UUCUCUUUUUCACUCUCUAA	2768	2787
		EIF4EBP2	CUUUCUUUCUUUUUUUCUGA	4515	4534
		HOXC13	CUCUCCUCUCUCUCUCUCUAG	785	805
		AFF2	UUGUCUCUCUCUCUCUCUGG	6935	6954
		HIPK2	CUUUUUCUUUUUCUUUUUAA	1261	1280
		RAB3B	UUCUCUCUUUUUUUUUUUAA	2370	2389
		KRT6A	CUUCUUCUCUCUCUCUCUAU	145	164

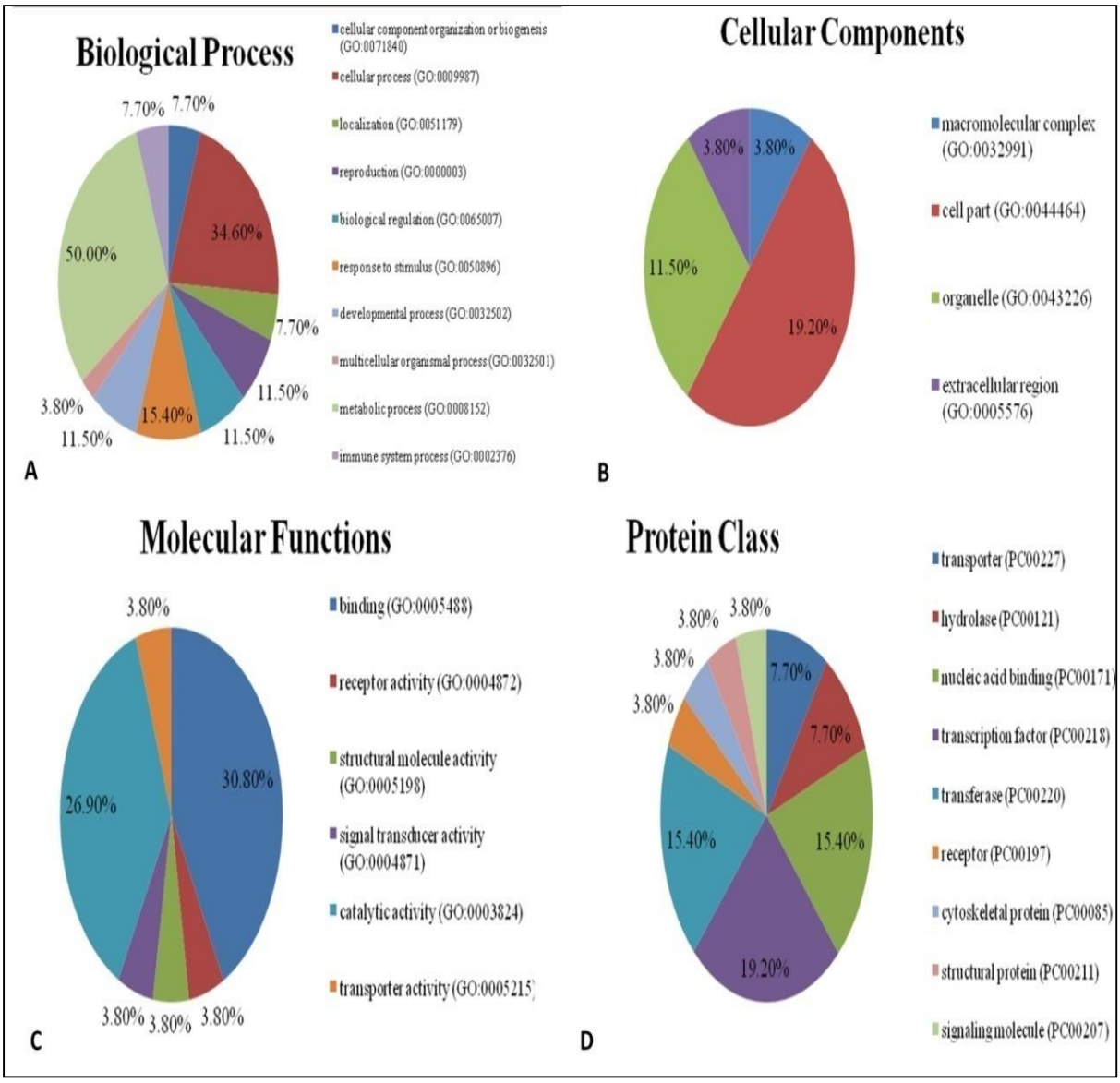

Figure 2: Gene ontology chart (A) Biological Process (B) Cellular components (C) Molecular functions (D) Protein class

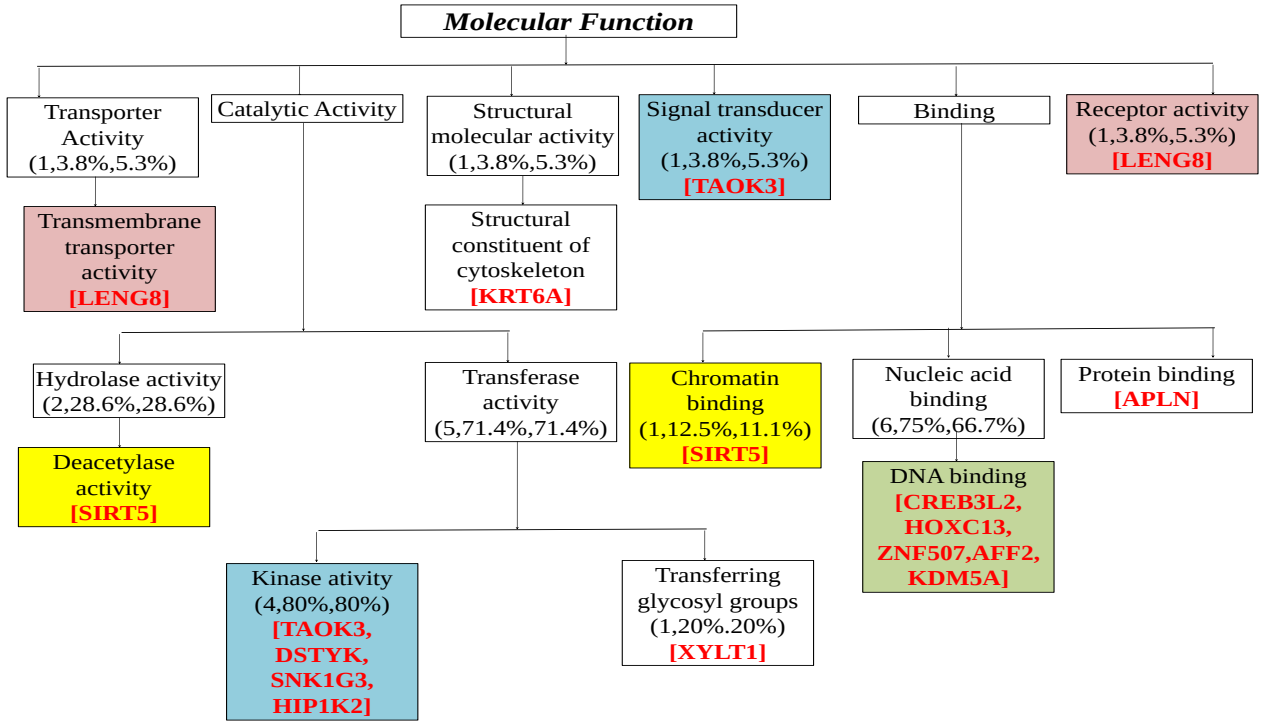

Figure 3: Involvement of potential human targets of predicted miRNAs of apricot in molecular function.

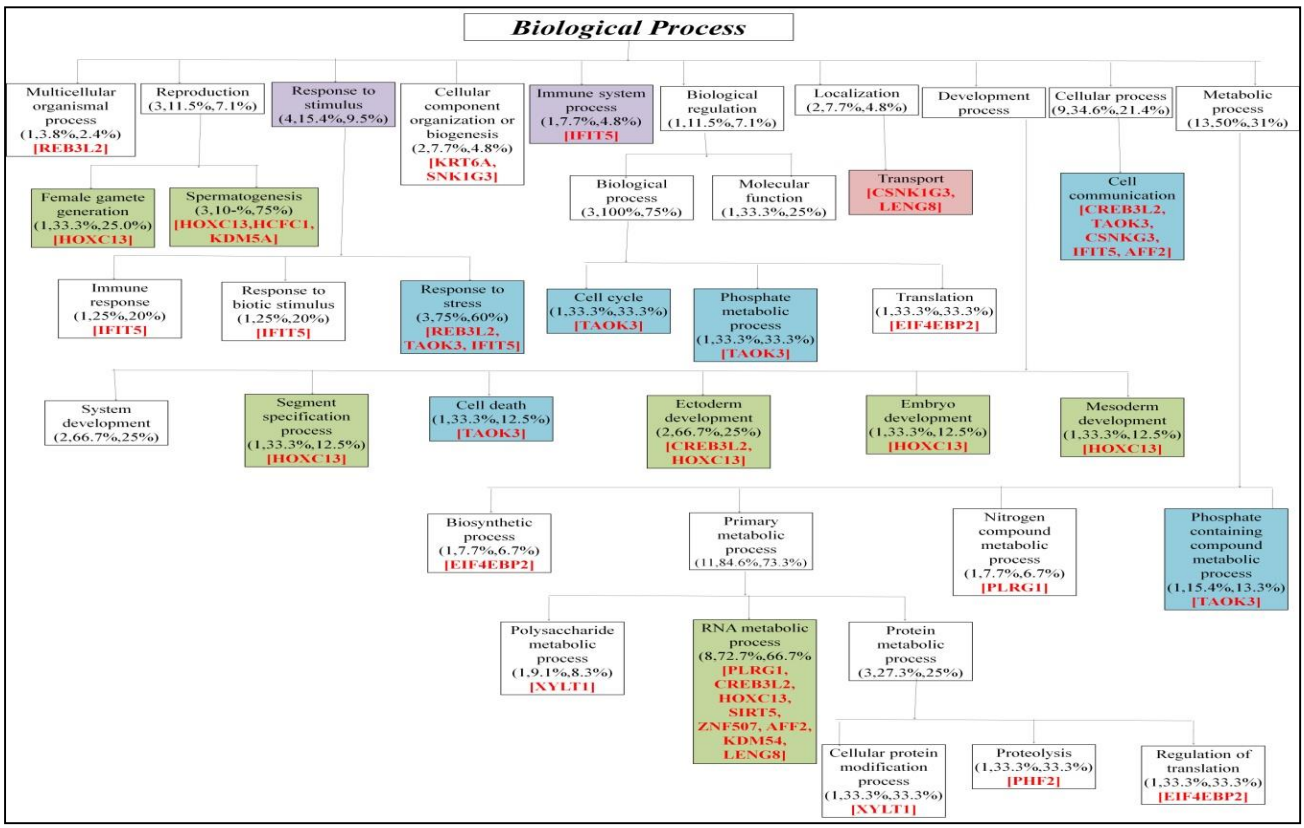

Figure 4: Involvement of potential human targets of predicted miRNAs of apricot in biological processes.

During gene ontology, we have observed that some of targets were found repetitively in multiple functions like SIRT5 involved in two molecular functions like chromatin binding and deacetylase activity [Figure 3] and LENG8 found in both biological process and molecular functions [Figure 3, 4]. We have also found transcription factors like HOXC13 and TAOK3 as targets of predicted miRNAs of apricots and are involved in cell process like cell death, cell-communication, cell cycle, etc. IFIT5 was found as target which is involved in regulation of immune system [Figure 4]. We have also noticed in pathway study, only two targets like CREM3L2 that are involved in Hetetrimeric G protein signaling pathway and WNT signaling pathway and second CSNK1G3 involved in Parkinson disease and transcription regulation of ZIP transcription factor [Table 4].

Table 4: Pathway analysis of human targets.

Target Name	Pathway
CREB3L2	Hetetrimeric G protein signaling pathway
	WNT signaling pathway
CSNK1G3	Parkinson disease
	Transcription regulation of ZIP transcription factor

Table 5: Functional and disease association of human targets.

Target Gene Identifiers	Gene Name	Category	Association
CREB3L2	CAMP Responsive Element Binding Protein 3 Like 2	Transcription activator	Myxofibrosarcoma and Sarcoma.
CSNK1G3	Casein Kinase 1 Gamma 3	Serine/threonine protein kinase	Involve in WNT signaling pathway
APLN	Apelin	endogenous ligand for the G-protein-coupled APJ receptor	Syndrome Of Inappropriate Antidiuretic Hormone and Severe Pre-Eclampsia
PLRG1	Pleiotropic Regulator 1	Protein coding gene	Poikiloderma With Neutropenia
XYLT1	Xylosyltransferase 1	Transferase enzyme	Desbuquois Dysplasia 2 and Pseudoxanthoma Elasticum

Figure 5: Multiple sequence alignment of predicted miRNAs of apricot with set of reported human miRNAs of specific targets (A) APLN (B) CSNK1G3 (C) PLRG1 and (D) XYLT1.

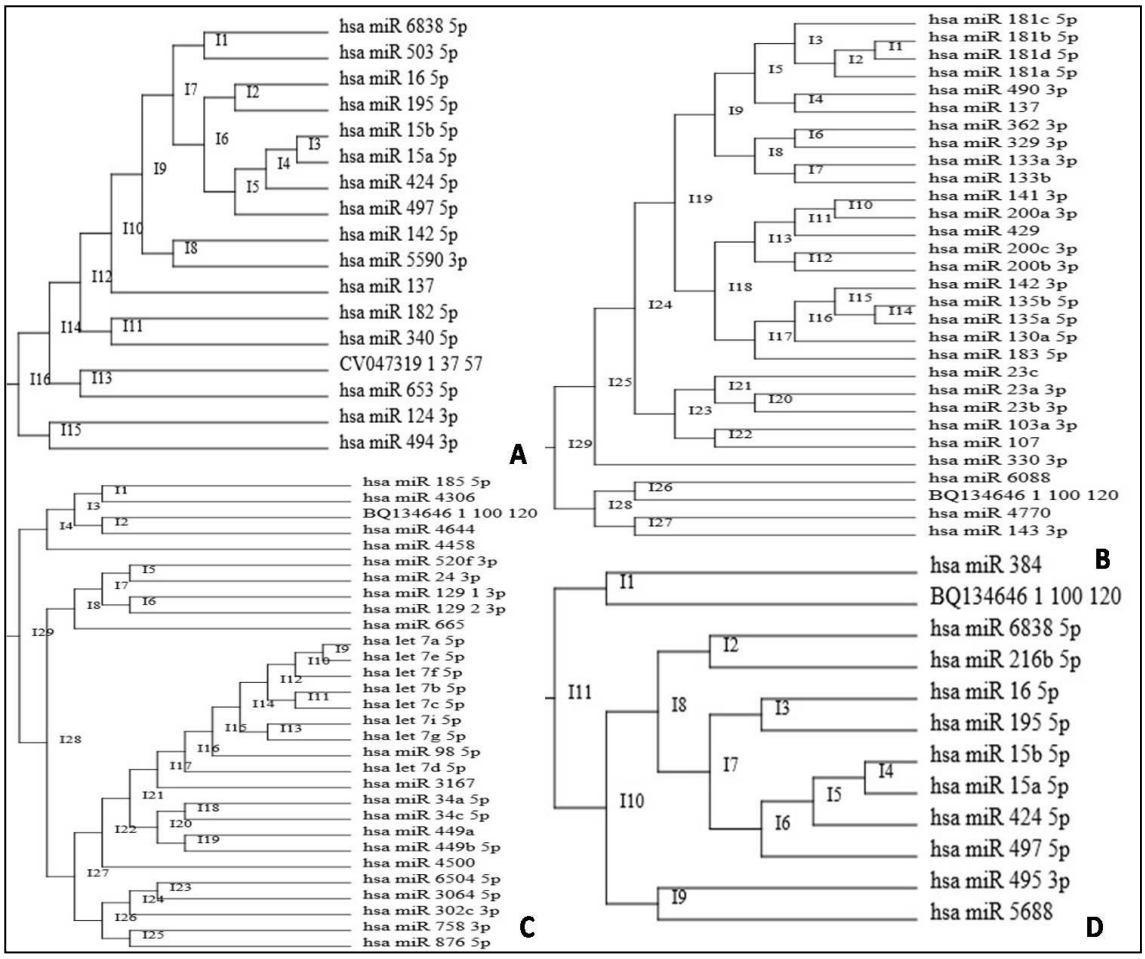

Figure 6: Secondary structure of pre-miRNAs of apricot

After functional annotation, we have performed multiple sequence alignment of predicted miRNAs of apricot with the set of reported human miRNAs which were derived from TargetScanHuman database for identify sequence similarity by using LocARNA tool. We have identified CV047319 have higher similarity with has-miR-653-5p which targets APLN and BQ134646 have higher similarity with has-miR-384 which targets PLRG1 [Figure 6]. On the other hand, human miRNAs of some targets i.e. PHF2, ZNF507, AFF2 and EIF4EBP2 have not found sequence similarity with predicted miRNAs of apricot. Therefore, we hypothesize that miRNAs like CV047319 and BQ1346464 of apricot might have higher achievability and potency to bind APLN, PLRG1, CSNK1G3 and XYLT1 and regulate their functions.

4. CONCLUSION

For non-model species, EST data provide valuable source to identify conserved miRNAs where genome sequence is not available. The prediction of miRNAs of *Prunus armeniaca* (Apricot) was improved by use of series of criteria focusing on miRNA sequence quality, stem-loop secondary structure and energy calculations. Based on sequence similarity with reported human miRNAsand complementarity with human targets, we have identified 2 (Two) potential miRNAs of apricot, which are reported for the first time computationally. The finding of this study will not only reinforce this bioinformatics approach to identify miRNAs of non-model species but also help in establishing cross-kingdom regulation of dietary miRNA on consumer's physiology.

CONFLICT OF INTEREST

The authors declare that no competing financial interests exist.

REFERENCES

1 Christopher AF, Kaur RP, Kaur G, Kaur A, Gupta V, Bansal P. MicroRNA therapeutics: Discovering novel targets and developing specific therapy. Perspect Clin Res. 2016 Apr-Jun;7(2):68-74. doi: 10.4103/2229-3485.179431. Review. PubMed PMID: 27141472; PubMed Central PMCID: PMC4840794.

2 Macfarlane LA, Murphy PR. MicroRNA: Biogenesis, Function and Role in Cancer.Curr Genomics. 2010 Nov;11(7):537-61. doi: 10.2174/138920210793175895. PubMed PMID: 21532838; PubMed Central PMCID: PMC3048316.

3: Zhang L, Hou D, Chen X, Li D, Zhu L, Zhang Y, Li J, Bian Z, Liang X, Cai X,Yin Y, Wang C, Zhang T, Zhu D, Zhang D, Xu J, Chen Q, Ba Y, Liu J, Wang Q, Chen J, Wang J, Wang M, Zhang Q, Zhang J, Zen K, Zhang CY. Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. Cell Res. 2012 Jan;22(1):107-26. doi: 10.1038/cr.2011.158. Erratum in: Cell Res. 2012 Jan;22(1):273-4. PubMed PMID: 21931358; PubMed Central PMCID: PMC3351925.

4 Pastrello C, Tsay M, McQuaid R, Abovsky M, Pasini E, Shirdel E, Angeli M, Tokar T, Jamnik J, Kotlyar M, Jurisicova A, Kotsopoulos J, El-Sohemy A, Jurisica I. Circulating plant miRNAs can regulate human gene expression in vitro. Sci Rep. 2016 Sep 8;6:32773. doi: 10.1038/srep32773. PubMed PMID: 27604570; PubMed Central PMCID: PMC5015063.

5 Minaiyan M, Ghannadi A, Asadi M, Etemad M, Mahzouni P. Anti-inflammatory effect of Prunus armeniaca L. (Apricot) extracts ameliorates TNBS-induced ulcerative colitis in rats. Res

Pharm Sci. 2014 Jul-Aug;9(4):225-31. PubMed PMID:25657793; PubMed Central PMCID: PMC4314870.

6 Boguski MS, Lowe TM, Tolstoshev CM. dbEST--database for "expressed sequence tags". Nat Genet. 1993 Aug;4(4):332-3. PubMed PMID: 8401577.

7 Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D140-4. PubMed PMID: 16381832; PubMed Central PMCID: PMC1347474.

8 Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003 Jul 1;31(13):3406-15. PubMed PMID: 12824337; PubMed Central PMCID: PMC169194.

9 Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. NucleicAcids Res. 2011 Jul;39(Web Server issue):W155-9. doi:10.1093/nar/gkr319. Epub2011 May 27. PubMed PMID: 21622958; PubMed Central PMCID: PMC3125753.

10 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP,Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25-9. PubMed PMID: 10802651; PubMed Central PMCID: PMC3037419.

11 Maruyama Y, Wakamatsu A, Kawamura Y, Kimura K, Yamamoto J, Nishikawa T, Kisu Y, Sugano S, Goshima N, Isogai T, Nomura N. Human Gene and Protein Database (HGPD): a novel database presenting a large quantity of experiment-based results in human proteomics. Nucleic Acids Res. 2009 Jan; 37(Database issue):D762-6. doi: 10.1093/nar/gkn872. PubMed PMID: 19073703; PubMed Central PMCID: PMC2686585.

12 http://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp

13 http://www.targetscan.org/vert_71/