BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net **Hypothesis**

Volume 8(2)

Identification of miRNAs in *C. roseus* and their potential targets

Ankur Omer[1]*, Sandeep Singh[2], Naveen Duhan[3]

[1]Assistant Professor, Department of Bioinformatics and Computers, Dolphin College of Life Sciences, Fatehgarh Sahib, Punjab, Pin Code: 140307, India; [2]Assistant Professor, Department of Biotechnology, Lovely Professional University (LPU), Jalandhar - Delhi G.T. Road (NH-1), Phagwara, Punjab, Pin Code: 144402, India; [3]Department of Bioinformatics and Computers, Dolphin College of Life Sciences, Fatehgarh Sahib, Punjab, Pin Code: 140307, India; Ankur Omer –Email: **:** ankuromer05@rediffmail.com; Phone: +91-8054343166; *Corresponding author

Received January 03, 2012; Accepted January 07, 2012; Published January 20, 2012

Abstract:
MicroRNAs are small (20-22 nucleotides) none coding, regulatory RNAs, whose pivotal role in gene expression has been associated in number of diseases, therefore prediction of miRNA is an essential yet challenging field. In this study miRNAs of *C. roseus* are predicted along with their possible target genes. A total of 19,899 ESTs were downloaded from dbEST database and processed and trimmed through SeqClean. Nine sequences were trashed and 31 sequences were trimmed by the program and the resulting sequences were submitted to Repeatmasker and TGICL for clustering and assembly. This contig database was now used to find the putative miRNAs by performing a local BLAST with the miRNAs of *B. rapa* retrieved from miRBase. The targets were scanned by hybridizing screened ESTs with the UTRs of human using miRanda software. Finally, 7 putative miRNAs were found to hybridize with the various targets of signal transduction and apoptosis that may play significant role in preventing diseases like Leukemia, Arthritis and Alzheimer.

Background:
MicroRNAs (miRNAs) are a large family of ~22 nucleotide long endogenous non-coding RNAs that play important posttranscriptional regulatory role in diverse organisms **[1].** Mature and functional miRNA have characteristic secondary structure (hairpin like) and develops from palindrome repeats **[2].** There are a number of methods for predicting miRNAs such as biochemical method, direct cloning, sequencing method and computational method. Apparently a distinct advantage of computational method is to identify even those miRNAs which are expressed only at a certain developmental stage of a cell that was not possible with other approaches of sequencing and cloning **[3].** The computational method of miRNA prediction depends largely upon the conservedness of miRNA sequences from organism to organism, that paved a strong approach of homology search method for predicting potential miRNAs. One of the most effective and rapid computational method of miRNA prediction is by EST analysis **[4].** *C. roseus* is a medicinal plant containing a number of alkaloids like Vincristin and Vinblastin that are effective against Leukemia **[5].** Plant extract has also been successfully used to develop effective drug against Diabetes mellitus **[6].** Since the first discovery of the miRNAs, miRNA abundance in the genomes of various organisms **[7]** including insects, plants **[8],** viruses and higher vertebrates has been reported **[9].** The miRNAs of closely related species and to some extent of distant species are generally present as conserved sequences **[10].** miRNAs acts as a regulatory element in various cellular and developmental processes *viz.* cell division **[11],** cell death **[12],** hormone secretion **[13]** and neural development **[14].** Recent evidences suggest that miRNAs also functions as tumour suppressors and oncogenes **[15].** They operate by binding to the 3'UTR region of an mRNA sequence with antisense base pairing and cleave the target mRNA or strangle its translation into protein. One miRNA can suppress various different mRNAs and a single mRNA may be bound by several assisting miRNAs **[16].** A number of approaches have been used to predict miRNAs in various organisms. Initially, miRNAs were identified using

 © 2012 Biomedical Informatics

genetic or biochemical methods **[17].** Later, direct cloning and sequencing of total small RNAs with appropriate size from isolated tissues or whole organisms enabled the identification of hundreds of miRNAs in plants and animals **[18].** In majority of recent experiments for miRNA prediction, computational approaches were firstly used to predict miRNAs followed by molecular techniques such as Northern blotting for validation **[19].** Evidently, computational approaches have played a progressively important role in miRNA identification. A discrete advantage of computational approaches is that the miRNAs which are expressed in specific tissues, at destined levels of development or at less copy number, can be readily identified by computational searching, whereas the approaches such as cloning and sequencing finds difficulty and often miss to identify miRNAs **[20].** The principles of computational approaches are based on the major characteristic features of microRNAs: hairpin-shaped stem-loop secondary structure with minimal folding free energy **[21]** and high evolutionary conservation from species to species **[22].** Accumulating evidence shows that many miRNAs are evolutionarily conserved in animals from worms to humans **[10],** suggesting a powerful strategy to predict potential miRNAs by using homology search. In fact, homology search as a computational approach has been developed to identify miRNA genes in both plants and animals **[23].**

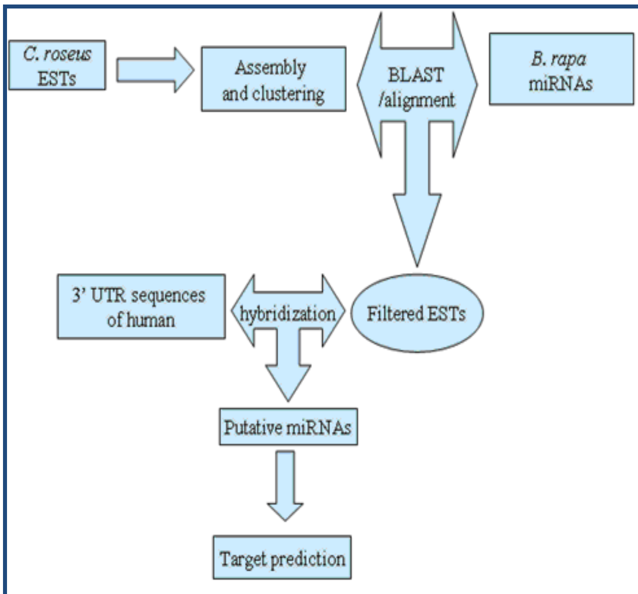


Figure 1: An overview of the used methodology during miRNA prediction of *C. roseus*.

Methods

C. roseus EST dataset:

As of April 27, 2011 dbEST [database of Expressed Sequence Tags]; contained 69,231,200 in the database. Among that there are 19,899 EST's of *Catharanthus roseus*, which were downloaded from the dbEST database of NCBI. The ESTs database has the highest set of impurities associated with them of 1.63%, which has to be removed before the further processing **[24].** SeqClean software was used to trim out the redundant sequences and removing the impurities. SeqClean **[24]** utilizes BLAST **[25]** to remove sequence highly similar (by default minimum 94% identity) to a given list of vectors, linkers, adaptors, or primer sequences that is located within 30% of total EST from the 3' or 5' end of the sequences to be trimmed **[26].** SeqClean also removes polyA repeats and applies low complexity filtering (in addition to performing sequence alignment) to identify similar vector segments in the target EST **[24].** The cleaned and trimmed ESTs obtained from the SeqClean were then processed via Repeatmasker (for masking the repeated sequences) and TGICL (for clustering) **[27].**

microRNA sequence dataset

miRBase is a database of validated and published miRNAs. miRNAs of *B. rapa* (closest homolog of *C. roseus* whose experimentally validated miRNAs were present) were downloaded from the miRBase.

Searching miRNA homologs and their secondary structure prediction

Each processed and clustered EST contigs obtained as output of TGICL were used to BLAST against the miRNAs dataset of *B. rapa* to identify ESTs having at least 60% identity. Secondary structures of resulting ESTs were calculated and their MFE (minimal folding energy) scores were recorded **[28].**

Hybridization using miRanda

The non redundant human UTRs were downloaded from UTRdb database. The targets for the putative miRNAs were then scanned in the 3' UTRs with the help of miRanda software. Hybridization energy as well as percentage identity were calculated for each EST with the help of Vienna RNA package **[29].** ESTs having good hybridization energy along with all the previous criteria were finally considered as putative miRNAs of *C. roseus.* All the possible targets genes for each putative miRNAs were predicted with the help of KEGG database of NCBI and their function was studied. **(Figure 1)** describes flowchart of the steps followed for the identification of miRNAs in *C. roseus*

Discussion:

ESTs refinement

Out of total 19,899 ESTs, 31 ESTs were trashed and 5,963 sequences were trimmed out resulting into a total of 19,868 ESTs. Now, the output obtained from SeqClean was processed as an input for Repeatmasker that masked the repeating sequences. Out of 10196333 bp (base pairs) 93026 bp got masked means only 0.91% of bp got masked and the result was found satisfactory as it clears the cut off range. Similar procedure was used for refining the contaminant vector sequences present in ESTs by Chen Y.A. *et al.* **[26].**

The matured and validated miRNAs of *B. rapa* were downloaded from the miRBase database to predict the putative miRNAs by local BLAST in local contig database of *C. roseus*. Ueno S. *et al.* used similar activity for clustering and assembly Oak tree ESTs **[30].**

Prediction of putative miRNAs

The miRNAs of *Brassica rapa* were compared with the assembled ESTs of *Catharanthus roseus* to identify regions in the Helianthus genome where experimentally validated miRNAs of *B. rapa* shows its presence in the contigs of *C. roseus*. Both mature and precursor miRNA matches were checked out in the resultant clustered contigs and singletons. This step generated

© 2012 Biomedical Informatics

filtered ESTs based on score and identity. **Table 1 (see supplementary material)** shows MFE (Minimum Folding Energy) value of 10 best putative miRNAs. The query that whether the miRNAs are blocking or repressing any of the human genes was solved by hybridizing putative miRNAs with the 3' UTRs of human through miRanda software. Peng G *et al.* used miRanda to find the regulating activity of let 7 miRNA during apoptosis **[31].** 7 out of 10 putative miRNAs were found to hybridise with the potential targets with a probability to block the action of those genes. **Figure 2** depicting the result of local BLAST between the miRNA and UTR sequences, minimum identity (63%) was found with the cath-7 miRNA and its related UTR. **Table 2 (see supplementary material**) shows the comparative account of finally selected 7 miRNAs on the basis of hybridization energy, Identity, Maximum score and the gene getting blocked. miRanda was used in order to hybridize the human UTRs with the filtered and screened ESTS dataset ultimately leading to the generation of 10 putative miRNAs, that were again filtered on the basis of the possible ability to silence the genes, traced back with the help of KEGG brite pathway.

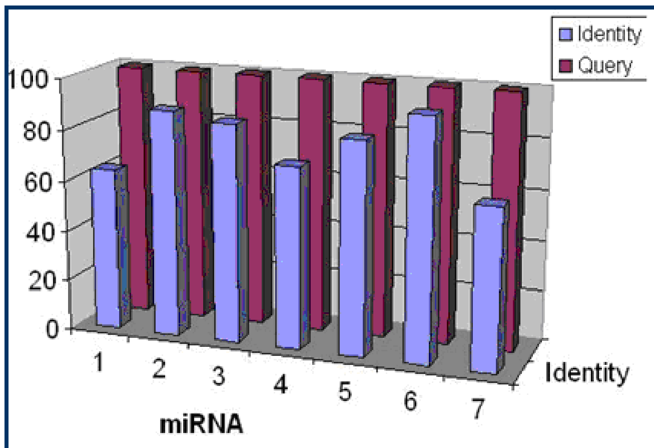


Figure 2: Showing the percentage of Identity between the predicted miRNAs and Query sequence possibly found to be involved in silencing particular target gene

Cath-miR1 was found to hybridize with the IL15RA and MGC104179 genes responsible for enhancing cell proliferation and expression of apoptosis inhibitor BCL2L1/BCL2-XL and BCL2 **[32].** cath-miR 2 hybridizes with CHRM1, HM1, M1R and MGC30125 genes that are responsible for mediating slow excitory post synaptic potential at postganglionic nerve **[33].** PARVG gene hybridized by cath-miRNA 3 is responsible for coding Parvin protein (are actin binding protein found to be associated with focal contacts) **[34],** cath miRNA 4 was found to be responsible for silencing various signal transducting and apoptotic genes. BMPR1B, ALK-6 and CDw293 genes hybridized by cath miRNA 5 encodes BMPR1B receptors involved in forming phalenges **[35].** Cath miRNA 6 hybridization products *i.e.* CDKL2, KKIAMRE and P56 encodes an enzyme Cyclin dependent Kinase like 2 related with ser/thr protein kinase activity **[36].** Cath-miR7 hybridizes with PIP5K1A (type I phosphatidylinositol-4-phosphate 5-kinase alpha protein) take part in synthesis of phosphatidylinositol-4,5-bisphosphate and related functions **[37].** The secondary structures and free energy values of the final 7 miRNAs are shown in **Table 3 (see supplementary material)**

Conclusion:

The *C. roseus* is a very important medicinal plant for human against various diseases. The miRNAs predicted from it were found to target various genes related with apoptotic, signal transduction and CDK pathways, therefore, they may prove helpful for either curing or preventing various diseases related to the genes they are targeting. The final predicted miRNAs may surve as potential resources to initiate experimental validation that may provide valuable information for functional analysis of miRNA regulated pathways.

Acknowledgements:

I would like to thank management of our college and colleagues for providing me the environment and necessary support to finish my work.

References:

[1] Khvorova A *et al. Cell* 2003 **115:** 209 [PMID: 14567918].
[2] Qiang G *et al. Chinese Science Bulletin.* 2007 **52:** 2380.
[3] Mishra AK & Lobiyal DK *Inter. J of Comp. Theory and Engineering* 2010 **2:**1793.
[4] Zhang Y *Nucleic Acids Res.* 2005 **33:**W701 [PMID: 15980567].
[5] El-Sayed A & Cordell GA *J Nat Prod.* 1981 **44 3:** 289 [PMID: 7264679].
[6] Ghosh RK & Gupta I *Indian J Animal Health* 1980 **19:**145.
[7] Lau NC *et al. Science* 2001 **294:** 858 [PMID: 11679671].
[8] Bartel B & Bartel DP. *Plant Physiol.* 2003 **132 :**709 [PMID:12805599].
[9] Lim LP *et al. Genes Dev.* 2003 **17:** 991 [PMID: 12672692].
[10] Reinhart BJ. *Nature* 2000 **403:** 901 [PMID: 10706289].
[11] Leaman D *et al. Cell* 2005 **121:**1097 [PMID: 15989958].
[12] Stark A *et al. PLoS Biol.* 2003 **1:** E60 [PMID: 14691535].
[13] Poy MN *et al. Nature* 2004 **432:** 226 [PMID: 15538371].
[14] Jin P *et al. Nat Neurosci.* 2004 **7:** 113 [PMID: 14703574].
[15] Calin GA *et al. Proc Natl Acad Sci.* 2002 **99:** 15524 [PMID: 12434020].
[16] Lewis B *et al. Cell* 2003 **115:** 787 [PMID: 14697198].
[17] Lee RC *et al. Cell* 1993 **75:** 843 [PMID: 8252621].
[18] Reinhart BJ *et al. Nature* 2000 **403:** 901 [PMID: 10706289].
[19] Zhang B *et al. Comput Biol Chem.* 2006 **30:** 395 [PMID: 17123865].
[20] Tang GQ & Maxwell ES *Genome Res.* 2008 **18:**104 [PMID: 18032731].
[21] Lagos-Quintana M *et al. Science* 2011 **294:**853 [PMID: 11679670].
[22] Grad Y *et al. Mol Cell* 2003 **11:**1253 [PMID: 12769849].
[23] Weber MJ *FEBS J.* 2005 **272:** 59 [PMID: 15634332].
[24] Falgueras J *et al. BMC Bioinformatics.* 2010 **11:** 38 [PMID: 20089148].
[25] Altschul SF *et al. J Mol Biol.* 1990 **215**: 403 [PMID: 2231712].
[26] Chen YA *et al. BMC Genomics.* 2007 **8:** 416 [PMID: 17997864].
[27] Smit AF. *Curr Opin Genet Dev.* 1996 **6:** 743 [PMID: 8994846].
[28] Stark A *et al. Genome Res.* 2007 **17:** 1865 [PMID: 17989255].
[29] Krek A *et al. Nat Genet.* 2005 **37:** 495 [PMID: 15806104].
[30] Ueno S *et al. BMC Genomics* 2010 **23**: 650 [PMID: 21092232**].**
[31] Peng G *et al. Brain Res Bull.* 2011 **86:** 272 [PMID: 21827835].

 © 2012 Biomedical Informatics

[32] Sato N *et al*. *Blood*. 2011 **117:** 4032 [PMID: 21304101].
[33] Messer WS *et al*. *Pharm Acta Helv*. 2000 **74:** 135 [PMID: 10812950].
[34] Yoshimi R_ *et al*. *J Immunol*. 2006 **176:** 3611 [PMID:16517730].
[35] Mishina Y *et al*. *J Biol Chem*. 2004 **279:** 27560 [PMID: 15090551].
[36] Marracci GH *et al*. *Biochem Biophys Res Commun*. 2006 **344:** 963 [PMID: 16631599].
[37] Kouchi Z *et al*. *Biochem Biophys Res Commun*. 2011 **408:** 523 [PMID: 21514270].

Edited by P Kangueane
Citation: Omer *et al*. Bioinformation 8(2): 075-080 (2012)
License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

BIOINFORMATION *open access*

Supplementary material:

Table 1: Showing MFE (Minimum Folding Energy) and sequence of all putative miRNAs

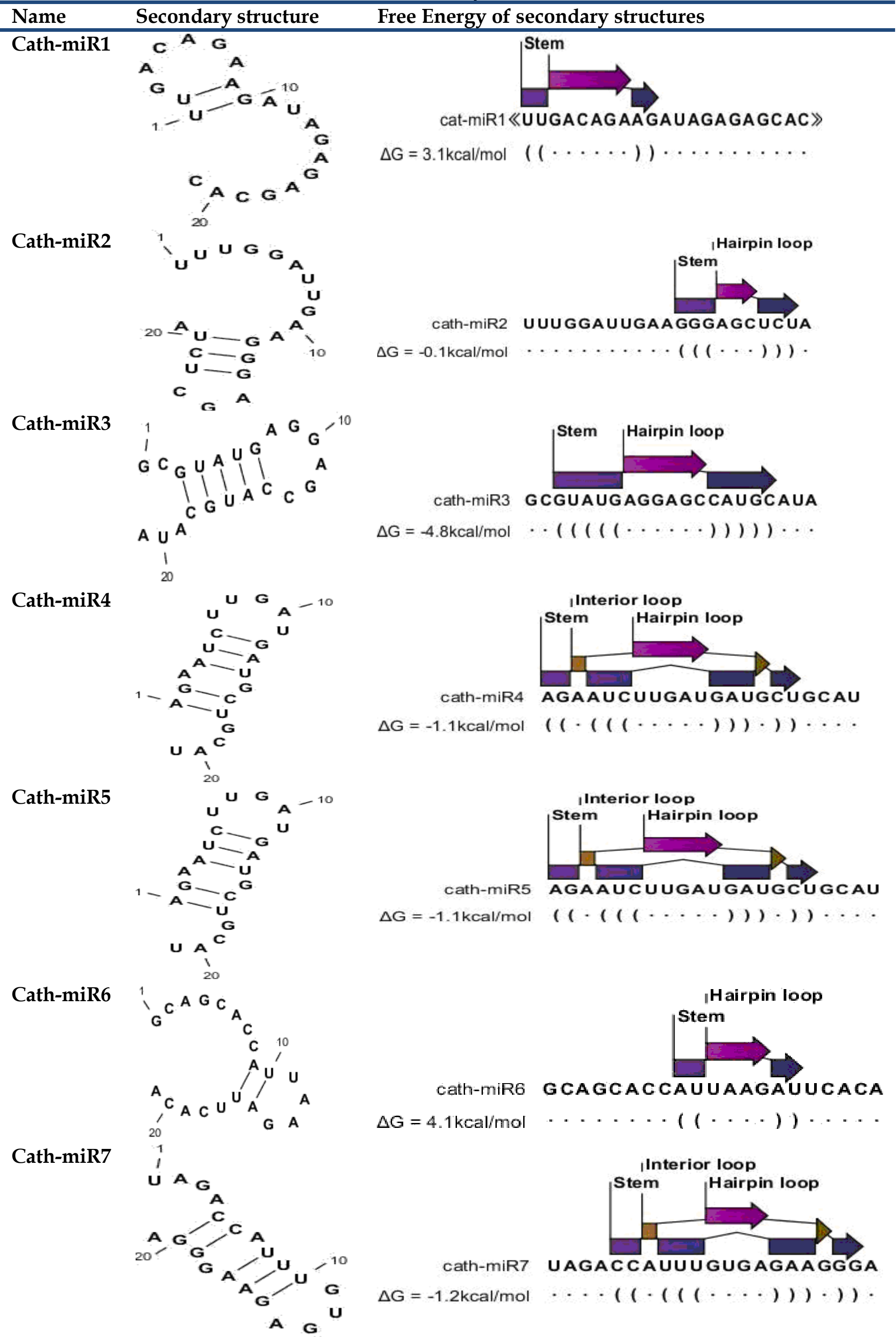
Name	MFE value	Predicted miRNA
>cath-miR1	3.1	UUGACAGAAGAUAGAGAGCAC
>cath-miR2	-0.1	UUUGGAUUGAAGGGAGCUCUA
>cath-miR3	-4.8	GCGUAUGAGGAGCCAUGCAUA
>cath-miR4	-1.1	AGAAUCUUGAUGAUGCUGCAU
>cath-miR5	-1.1	AGAAUCUUGAUGAUGCUGCAU
>cath-miR6	4.1	GCAGCACCAUUAAGAUUCACA
>cath-miR7	-1.2	UAGACCAUUUGUGAGAAGGGA
>cath-miR8	0.0	CAUCAAUGAAAGGUAUGAUUCC
>cath-miR9	-0.8	UAAUCUGCAUCCUGAGGUUUA
>cath-miR10	-0.8	UAAUCUGCAUCCUGAGGUUUA

Table 2: The hybridization energy (cutoff score was taken to be -20 kCal/mol), identity percentage and the gene getting silenced due to respective miRNA are shown.

Name	Hybridisation Energy	Identity percentage	Score	Gene blocked
>cath-miR1	-21 kCal/mol	64	155	IL15RA and MGC104179
>cath-miR2	-25.5 kCal/mol	88.9	175	CHRM1, HM1, M1R & MGC30125
>cath-miR3	-28.14 kCal/mol	85.71	175	PARVG
>cath-miR4	-22 kCal/mol	71.4	161	Apoptotic & Signal transducting
>cath-miR5	-20.31 kCal/mol	83.34	171	BMPR1B, ALK-6 and CDw293
>cath-miR6	-24.5 kCal/mol	94.44	171	CDKL2, KKIAMRE and P56
>cath-miR7	-22.19 kCal/mol	63	160	PIP5K1A
>cath-miR8		NO HYBRIDISATION		
>cath-miR9		NO HYBRIDISATION		
>cath-miR10		NO HYBRIDISATION		

© 2012 Biomedical Informatics

Table 3: Illustration of miRNA structures: Secondary structures of the final selected 7 miRNAs along with their free energy values



 © 2012 Biomedical Informatics