Computational identification of microRNAs and their targets in *Catharanthus roseus* expressed sequence tags

Alok Pani, Rajani Kanta Mahapatra *

School of Biotechnology, Campus-11, KIIT University, Bhubaneswar, Odisha 751024, India

**A R T I C L E   I N F O**

Article history:
Received 31 March 2013
Accepted 28 June 2013
Available online 14 August 2013

Keywords:
Vincristine
miRNAs
ESTs
Computational prediction
*C. roseus*

**A B S T R A C T**

No study has been performed on identifying microRNAs (miRNAs) and their targets in the medicinal plant, *Catharanthus roseus*. In the present study, using the comparative genomics approach, we have predicted two potential *C. roseus* miRNAs. Furthermore, twelve potential miRNA targets were identified in *C. roseus* genome based on the characteristics that miRNAs exhibit perfect or nearly perfect complementarity with their targeted miRNA sequences. Among them many of the targets were predicted to encode enzymes that regulate the biosynthesis of terpenoid indole alkaloids (TIA). In addition, most of the predicted targets were the gene coding for transcription factors which are mainly involved in cell growth and development, signaling and metabolism. This is the first in silico study to indicate that miRNA target gene encoding enzymes involved in vinblastine and vincristine biosynthesis, which may help to understand the miRNA-mediated regulation of TIA alkaloid biosynthesis in *C. roseus*.

© 2013 The Authors. Published by Elsevier Inc. Open access under CC BY-NC-ND license.

1. Introduction

The medicinal plant *Catharanthus roseus* L. G. Don is of enormous pharmaceutical potential because of the presence of >120 terpenoid indole alkaloids (TIA), some of which are known to exhibit strong pharmacological activities [1]. Vinblastine and vincristine, the antineoplastic bisindole alkaloids are produced only in trace amounts; vinblastine (0.01%) and vincristine (0.003%). Furthermore, with significant international efforts cell cultures are also not yet a valid alternative for production for these low yielding secondary metabolites [2]. Therefore, a deeper understanding of the regulatory system governing TIA metabolism is of particular interest and could eventually make successful metabolic engineering of alkaloid biosynthesis possible.

Several hundred miRNAs have been identified in plants by computational and experimental approaches [3,4]. However, a little is known about experimental or computational identification of miRNA in *C. roseus* species. *C. roseus* belongs to Apocynaceae family, reflecting a disparity between the important values of this plant family and insufficient molecular and genetic studies, including small RNA mediated gene regulation. So to gain insight into miRNAs and their important regulatory functions in terpenoid indole alkaloid biosynthetic pathway, we studied miRNA and their targets in *C. roseus* genome using computational approach.

2. Materials & methods

2.1. miRNA reference set

To search potential miRNAs, a total of previously known 328 miRNAs from *Arabidopsis thaliana* were obtained from miRNA Registry Database (Release 18.0, November 2011). Although some of these *A. thaliana* miRNAs were initially identified by computational approaches, a majority of them have been validated by experimental approaches including direct cloning, PCR, Northern blotting, and/or 5′ rapid amplification of cDNA end (5′RACE) [3]. To avoid the overlap of miRNAs, the redundant miRNA sequences were removed manually and the remaining sequences were used as a reference miRNA for homologous prediction in *C. roseus*. We have referred to the previous work on computational prediction of miRNAs [5].

2.2. Availability of software

Comparative software BLAST-2.2.14 was used from NCBI Genbank. MFOLD 3.1 was used online to analyze secondary structure of RNAs. MirEval (http://tagc.univ-mrs.fr/mireval) was used to predict miRNA precursor [6]. These precursor sequences were used for BLASTx analysis for removing the protein-coding sequences and retained only non-protein encoding sequences. BLASTn from NCBI (http://www.ncbi.nlm.nih.gov) was used to analyze potential targets of miRNAs.
2.3. Prediction of miRNAs

Procedure for searching potential miRNAs in *C. roseus* is shown in Fig. 1. The non-redundant miRNA sequences after initial screening were subjected to BLASTn search for *C. roseus* miRNA homolog against EST database. The adjusted BLASTn parameters setting were as follows: expect value was set at 1000 to increase the hit chance for more potential sequences; low complexity was chosen as the sequence filter; the number of descriptions and alignments were raised to 1000. The default word match size between the query and a database sequence was 7. RNA sequences were considered as miRNA candidates only if they fit the following criteria: (1) at least 18 nt length were adopted between the predicted and mature miRNAs and (2) allowed to have 0–3 nt mismatches in sequence with all previously known plant mature miRNAs. The ESTs that closely match the previously known plant mature miRNAs were included in the set of miRNA candidates and used for additional characterization by subjected to evaluation for miRNA precursor prediction properties using mirEval software. These precursor sequences were used for BLASTx analysis for removing the protein-coding sequences and retaining only the non-protein-coding sequences.

2.4. Prediction of secondary structure

Precursor sequences of these potential miRNA homologs were subjected to hairpin structure prediction using the Zuker folding algorithm with MFOLD 3.1 [7]. In brief, the following criteria were applied in designating the RNA sequence as a miRNA homolog described by (1) the sequence could fold into an appropriate stem-loop hairpin secondary structure; (2) the small RNA sits in one arm of the hairpin structure; (3) no more than 6 mismatches are between the predicted mature miRNA sequence and its opposite miRNA (miRNA*) sequence in the secondary structure; (4) no loop or break is in the miRNA or miRNA* sequences, and (5) predicted secondary structure has higher MFEI and negative MFE [8].

The MFEI was calculated using the following equation:

\[
\text{MFEI} = \left(\frac{\text{MFE}}{\text{length of the RNA sequence}}\right) \times 100 / (\text{G} + \text{C}) \%
\]

MFE denotes the negative folding free energies (ΔG).

2.5. Prediction of mRNA targets of miRNAs

In brief, we used the potential *C. roseus* miRNA BLAST analysis against the *C. roseus* mRNA database to search sequences conforming to the following standards as the *C. roseus* candidate targets gene. (1) The maximum number of mismatched nucleotides between the mature miRNA and its potential target genes was four; (2) the maximum number of mismatched nucleotides at positions 1–9 was one; (3) no mismatches was allowed at positions 10–11; (4) more than two continuous mismatches at any position were not allowed [9].

2.6. Analysis of GO and KEGG pathway

To better understand the function of *C. roseus* miRNAs, Blast2go [10,11] was employed to investigate the predicted target genes. First, the identified miRNA targeted mRNAs were used to BLASTX against NR database. Second, the best hits identified by BLASTX were further searched against the GO and KEGG databases using default settings.

![Fig. 1](image-url). Procedure of potential *C. roseus* miRNA gene search by identifying homolog’s of previously known *A. thaliana* miRNAs.
3. Results

3.1. Potential miRNAs in C. roseus

Following the procedure depicted in Fig. 1, 19,910 ESTs from C. roseus were searched against 321 mature miRNA sequences of Arabidopsis. In total, two potential miRNAs were predicted in C. roseus. The two identified C. roseus candidate miRNAs belong to miR5021 family. The two predicted microRNAs are having maximal mismatches of 2 & 3 respectively against its homolog (Fig. 2). The length of the two EST is 611 nt and 592 nt while the precursor’s length is 80 nt and 100 nt respectively as identified by MirEval software. The mature miRNA sequences are located at the 3’ end of the miRNA precursors. Minimal folding free energy (MFE) is an important characteristic that determines the secondary structure of nucleic acids (DNA and RNA). The lower the MFE is, the higher the thermodynamically stable secondary structure of the corresponding sequence is [12]. The MFE value of the structures is estimated to $-17.90 \text{ kcal/mol}$ and $-34.70 \text{ kcal/mol}$ respectively. We have also calculated the MFE index (MFEI) for each sequence, to avoid false calling of other RNAs as miRNA candidates. The MFE index (MFEI) for each sequence was calculated as previously reported [13]. In this study, the MFEI value is 0.57 and 0.83 respectively. During the screening of the potential miRNAs, candidate miRNAs were evaluated for A + U content. The sequences of the miRNA precursors have A + U content of 61.25% and 58% respectively (Table 1), which is in agreement with the previous results [5]. It is estimated that in plants, approximately 10,000 ESTs contain 1 miRNA [14]. Therefore, the total of 19,910 ESTs in C. roseus examined in this study may contain 1–2 miRNAs. C. roseus belongs to Apocyanaceae family. Unfortunately, not a single miRNA from Apocyanaceae family has been deposited in the MiRBase [15]. We expect that as more miRNAs of this plant family are publicly available, more miRNAs will be identified in C. roseus. ESTs are partially transcribed gene sequences, which have been used to confirm the existence and expression of potential miRNAs predicted by computational approaches in Arabidopsis, rice and maize [16,17]. In this study, we also tested the predicted C. roseus miRNAs individually against the EST databases of GenBank. Our BLASTn search results indicated that several predicted periwinkle miRNAs exist in C. roseus EST databases, suggesting that these miRNAs were expressed in the Catharanthus genome. Moreover, many of the reference set miRNAs from Arabidopsis which are used for homology search in C. roseus have been validated by experimental approaches. So the result from the computational prediction will be useful to guide experimental design for biological verification.

4. Discussion

4.1. C. roseus miRNA targets and their functions

The miRNA-regulated genes control a variety of biological and metabolic processes. Gaining insight into the miRNA targets will help us to understand the spectrum of miRNA regulation and elucidate the functional importance of miRNAs. Increasing evidences have demonstrated that most plant miRNAs bind to their target mRNA sequences with perfect or near-perfect sequence complementarity [18,19]. This provides a powerful strategy for discovering potential miRNA targets by comparing and aligning miRNAs with miRNAs sequences. Here, we performed more stringent criterion [19] to identify potential C. roseus targets. After a set of screening criteria as described in the method, we achieved 12 target genes. Out of these 12 target genes many they have...
Interestingly, another transcriptionally regulated strictosidine synthase condenses tryptamine synthase is another potential target of miR5021 in alkaloids as well as quinoline alkaloids. In addition, strictosidine and key intermediates in the biosynthesis of monoterpenoid indole alkaloids are predicted for mir5021 with least Expect (E) value. These predicted mRNA targets are statistically significant in compare to the other mRNA targets with least Expect (E) value.

Out of several mRNA targets predicted; one target gene identified for C. roseus miRNA (EG560894) is UDP-glucose iridoid glucosyltransferase. This enzyme is responsible for glucosylation step in the biosynthetic pathway of the secondary metabolite, iridoid in higher plants [20]. They are pharmacologically active principles in various medicinal plants and key intermediates in the biosynthesis of monoterpenoid indole alkaloids as well as quinoline alkaloids. In addition, strictosidine synthase is another potential target of mir5021 in C. roseus. The transcriptionally regulated strictosidine synthase condenses tryptamine and the iridoid secoligandin to yield strictosidine, the universal precursor of TIA [21]. Interestingly, another Arabidopsis mir5021 homolog in C. roseus is predicted to target GCPE protein. GCPE protein is involved in the terpenoid biosynthesis process [22].

It is expected that freeing the expression of key enzymatic activities form the strict regulation to which they are normally subjected is expected to increase the flux through the pathway and product formation. Therefore, the next major steps are to successfully apply metabolic engineering approach, to regulate the enzymes through miRNA mediated regulation. This will help to explore unconventional alternate strategies that are economically viable for the commercial production of indole alkaloids.

In addition EG560894 was also predicted to target type-A response regulator (EF625489), which is involved in the final steps of a histidine-to-aspartate phosphorolay in cytokinin (CK) signaling in C. roseus. CKs are plant growth regulators with pleiotropic functions in plants. They can also control some secondary metabolite biosynthetic pathways in Arabidopsis [23] and terpenoid indole alkaloids (TIAs) in periwinkle cells [24].

Recent studies have shown that miRNAs are also involved in plant adaptation to environmental stresses, such as cold [25,26], salt [27], drought [28], and nutrient deficiency [29,30]. Interestingly, we identified one target of mir-5021 is putative secretory peroxidase [EF661875] which response to oxidative stress. Further analysis of Gene ontology (GO) suggested that they are involved in peroxidase activity.

C. roseus miRNA also target a gene chloroplast geranylgeranyl diphosphate synthase (EF625539) that is involved in isoprenoid biosynthesis. Isoprenoids are a large and highly diverse group of natural products with many functions in plant primary and secondary metabolism [31]. Another Arabidopsis mir-5021 homolog of C. roseus, EG565843 targets ADP-riboisolation factor 1. It is a GTP binding protein involved in the regulation of vesicle-mediated protein transport through the secretory pathway [32].

Many miRNA targets identified by bioinformatics and/or experimental methods were transcription factors that help control plant growth and development. In this study we found a target of mir-5021 called MYB transcription factor. MYB proteins are key factors in regulatory networks controlling development, metabolism and responses to biotic and abiotic stresses [33]. In Arabidopsis, the other functions of MYB transcription factor are (i) primary and secondary metabolism, (ii) cell fate and identity, (iii) developmental processes and (iv) responses to biotic and abiotic stresses. Magnesium chelatase subunit H is also predicted to be a target for mir5021 homolog in C. roseus. Previous studies have shown that magnesium-protoporphyrin IX (ProtoIX) chelatase large subunit (Mg chelatase H subunit; CHLH) is an ABA receptor. It mediates ABA responses in seed germination, post germination growth, and stomatal movement [34]. CHLH has multiple functions in plant cells. As a subunit of the Mg-chelatase, CHLH catalyzes the introduction of Mg to ProtoIX, a key regulatory step of chlorophyll biosynthesis. In addition, CHLH plays a key role in mediating plastid-to-nucleus retrograde signaling [35].

### Table 1

| miRNA     | Reference species | Gene ID     | EST length (nt) | NM (nt) | LM (nt) | LP (nt) | Side | A + U (%) | MFE  | MFEI |
|-----------|-------------------|-------------|-----------------|---------|---------|---------|------|-----------|------|------|
| mir-5021  | A. thaliana       | EG560894    | 697             | 2       | 20      | 80      | 3'   | 61.25     | 17.90| 0.57 |
| mir-5021  | A. thaliana       | EG565843    | 795             | 3       | 20      | 100     | 3'   | 58        | 34.70| 0.83 |

Note: NM, number of mismatch; LM, length of mature miRNAs; LP, length of precursor; MFE, minimal folding free energy; MFEI, minimal folding free energy index.

### Table 2

| miRNA family | Target accession ID | Target description | Function |
|--------------|---------------------|--------------------|----------|
| mir-5021     | EF625593            | MYB transcription factor | Transcription factor |
|              | EF625552            | ADP-ribosylation factor 1 | Protein transport |
|              | EF625539            | Geranylgeranyl diphosphate synthase | Isoprenoid biosynthetic process |
|              | EF625523            | Gamma-tocopherol methyltransferase related protein | Methyltransferase activity |
|              | EF625416            | GCPE protein | Terpenoid biosynthetic process |
|              | EF625531            | UDP-glucose glucosyltransferase | Transferase activity |
|              | EF625513            | Cytochrome c oxidase subunit I | Oxidative phosphorylation |
|              | EF625489            | Type-A response regulator | Regulation of Transcription |
|              | EF625499            | UDP-glucose iridoid glucosyltransferase | Transferase activity |
|              | EF625448            | Magnesium chelatase subunit H | Biological process |
|              | EF625362            | Chloroplast terpenoid cyclase | Terepase synthase activity |
|              | EF661875            | Putative secretory peroxidase | Response to oxidative stress |

### Table 3

| miRNAs | Biological Process | Accession IDs for the targets | GOs |
|--------|--------------------|------------------------------|-----|
| mir-5021 | DNA binding | EF625593 | GO: 003677 |
|         | GTP binding     | EF625552 | GO: 0005525 |
|         | Geranyltransferase activity | EF625539 | GO: 004317 |
|         | Methyltransferase activity | EF625523 | GO: 0008168 |
|         | Iron ion binding | EF625416 | GO: 0005306 |
|         | Transferase activity | EF625531 | GO: 0000758 |
|         | Oxidative phosphorylation | EF625513 | GO: 0004129 |
|         | Regulation of transcription | EF625489 | GO: 0000156 |
|         | Transferase activity | EF625469 | GO: 0016758 |
|         | Biological process | EF625448 | GO: 0009038 |
|         | Terepase synthase activity | EF625362 | GO: 0010333 |
|         | Response to oxidative stress | EF661875 | GO: 0004601 |
5. Conclusion

C. roseus (periwinkle) is the sole source of antinecancerous alkaloids vinblastine and vincristine which today is widely used for treatment of cancer. Ajmalicine is used as an antihypertensive alkaloid. Recent report demonstrates that, in addition to its antinecancerous properties, the extracts from the leaves of this plant can be used as prophylactic agent in many infectious diseases. The unfortunately low yield of the antineoplastics bisanidine alkaloids and with the failure of the alternative production systems, for example by in vitro culture of C. roseus cells, an attempt can be made for the genetic manipulation of the alkaloid biosynthesis pathway for higher production levels. Our study is the first in silico study to identify miRNAs and their targets in C. roseus, which we hope could help to better understand miRNA-mediated regulation of genes related to terpenoid indole alkaloids biosynthesis.

Competing interests

The authors have declared that no competing interests exist.

Acknowledgments

AP and RKM thank Dr. Mrutyunjay Suar, Director of School of Biotechnology in KIIT University for his encouragement and support and the Bioinformatics lab facility of School of Biotechnology, KIIT University for his encouragement and support.

To further understand the function of C. roseus miRNAs, the predicted target miRNAs were subjected to analysis by GO and Kyoto Encyclopedia of Genes and Genomes (KEGG), a database for analyzing gene functions systematically [36]. The result suggested that C. roseus miRNAs were involved in various biological processes such as oxidation–reduction process, regulation of transcription, transport, growth and development, metabolism and translation (Table 3). Pathway enrichment analysis, based on the KEGG database, demonstrates that the identified miRNAs participated in four metabolic networks. These networks were involved in terpenoid backbone biosynthesis, protein processing, porphyrin and chlorophyll metabolism, oxidative phosphorylation, and other secondary metabolites biosynthesis process (Table 4). Obviously, our study will help further understanding of the important regulatory roles of miRNAs in C. roseus growth and development, stress response, and likewise in research and development to augment the production of vinblastine and vincristine biosynthesis.

Table 4

| miRNAs     | Accession ID for targets | Target description | Enzyme | Pathways               |
|------------|--------------------------|--------------------|--------|------------------------|
| mir-5021 (EC558543) | EF265416 | CGPE protein | EC: 17.1.7.1 | Terpenoid backbone biosynthesis |
|            | EF265539 | Geranylgeranyl diphosphate synthase, type II | EC: 2.5.1.1 | Terpenoid backbone biosynthesis |
|            | EF266531 | UDP-glucose:glycoprotein glycosyltransferase | EC: 2.4.1.1 | Protein processing in ER |
|            | EF265513 | Cytochrome c oxidase subunit 1 | EC: 1.9.3.1 | Oxidative phosphorylation |
|            | EF265448 | Magnesium chelatase subunit H | EC: 6.6.1.1 | Porphyrin and chlorophyll metabolism |

References

[1] R. Van der Heijden, D.J. Jacobs, W. Snoeijer, D. Hallard, R. Verpoorte, The Catharanthus alkaloids: pharmacognosy and biotechnology. Curr. Med. Chem. 11 (2004) 607–628.
[2] R.J. Noble, The discovery of the vinca alkaloids–chemotherapeutic agents against cancer. Biochem. Cell Biol. 68 (1990) 1344–1351.
[3] S. Griffiths-Jones, R.J. Grocock, S. van Dongen, A. Bateman, J. Enright, miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 34 (2006) D140–D144.
[4] B.H. Zhang, X.P. Pan, C.H. Cannon, G.P. Cobb, T.A. Anderson, Conservation and divergence of plant microRNAs genes. Plant J. 46 (2006) 243–259.
[5] B. Zhang, Q. Wang, K. Wang, X. Pan, F. Liu, T. Guo, G.P. Cobb, T.A. Anderson, Identification of cotton microRNAs and their targets. Gene 397 (2007) 26–37.
[6] W. Ritchie, F.X. Théodule, D. Gautheret, Mireval: a web tool for simple microRNA prediction in genome sequences. Bioinformatics 24 (11) (2008) 1304–1309.
[7] M. Zuke, Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 13 (2003) 3406–3415.
[8] L. Wang, H. Liu, D. Li, H. Chen, Identification and characterization of maize microRNAs involved in the very early stage of seed germination. BMC Genetics 12 (2011) 154.
[9] F.Xie, T.P. Frazier, B.Zhang, Identification and characterization of micro rRNA and their targets in the bioenergy plantswitchgrass (panicum virgatum). Planta 232 (2010) 417–434.
[10] A. Conesa, S. Gotz, J.M. Garcia-Gomez, J.Torel, M. Talon, M. Robles, Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (2005) 3674–3676.
[11] A. Conesa, S. Gotz, Blast2go: a comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics (2008) 619832.