DIFFERENTIAL EXPRESSION OF INDOLE ALKALOID PATHWAY GENES ACROSS CATHARANTHUS ROSEUS PLANT ORGANS

AHMED M. RAMADAN1,2, MUNA A. ABDULGADER1, THANA KHAN1, NOUR O. GADALLA1,3 AND AHMED BAFIELDIN1,5

1Department of Biological Science, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia
2Agricultural Genetic Engineering Research Institute (AGERI), Agriculture Research Center (ARC), Giza, Egypt
3Department of Arid Land Agriculture, Faculty of Meteorology, Environment and Arid Land Agriculture, King Abdulaziz University, Jeddah, Saudi Arabia
4Genetics and Cytology Department, Genetic Engineering and Biotechnology Division, National Research Center, Dokki, Egypt
5Department of Genetics, Faculty of Agriculture, Ain Shams University, Cairo, Egypt

Corresponding author’s email: aamara@kau.edu.sa

Abstract

Vinblastine and vincristine are two indole alkaloids of the medicinal plant Catharanthus roseus, widely used in cancer chemotherapy such as leukemia, Hodgkin's disease and solid tumors. Transcriptomic data of C. roseus across different organs (stem, hairy root, flower, immature and mature leaves) was collected and de novo assembled then differential expression analysis was achieved. The target of this study was the detection of components of terpenoid indole alkaloid pathway based on gene expression analysis. The in-silico analysis was focused on six genes (SLS, sgd STR, OMT, DAT and GHR of the TIA pathway that were expressed in different plant organs. The results indicated possibility of formed enzyme in organ and transfer to other organs to work and restore the TIA pathway. The highest expressions level of most TIA key genes took places in root, flowers and mature leaves.

Key words: Transcriptome analysis, vinblastine, vincristine, de novo, in silico.

Introduction

Plants are attractive sources of pharmaceutical compounds used for therapeutic purposes, especially anti-cancer compounds, albeit being naturally synthesized in plants for the defence against pests and diseases (Ziegler & Facchin, 2008; Roepke et al., 2010; Perveen et al., 2020). Several other important pharmaceutical compounds exist in higher plants such as baccharin, ellipticine, homoharringtonine, triptolide, bruceantin, indicine-N-oxide, thalicarpine and maytansine (Ziegler & Facchin, 2008; Guirimand et al., 2010). Among plant products that are marketed as anti-cancer drugs are vinblastine, vincristine, etoposide and camptothecin. Some of these pharmaceutical compounds, e.g., vinblastine and vincristine, are produced in terpenoid indole alkaloid (TIA) pathway of Catharanthus roseus and are widely used in cancer chemotherapy against various leukemia, Hodgkin's disease, solid tumors, and depression (Contin et al., 1998; Dogru et al., 2000; Lange et al., 2001; Burlat et al., 2004; Hamid et al., 2017).

Apocynaceae is a family of medicinal plants including about 550 genera that are considered as a rich resource of TIA. Catharanthus roseus is a member of this family and comprises about 130 TIA; 25 of them are dimeric indole alkaloids. TIA biosynthesis pathway has received considerable attention; however, the full characterization of alkaloids pathways is not completely accomplished (Zhu et al., 2015). TIA's are metabolites that have antitumor features such as vincristine and vinblastine that are utilized to treat many types of cancer (Sofowora et al., 2013; Greenwell and Rahman, 2015; Nejat et al., 2015; Soares et al., 2016). The general mechanisms of these metabolites are destroying cancer cells through inhibiting tubulin polymerization; and attacking enzymes involved in DNA duplication and transcription such as topoisomerases I and II (Chi et al., 2015; Chan et al., 2015; Montecucco et al., 2015).

Pharmaceutical compounds synthesized in C. roseus plant are extracted at low concentrations. The concentrations of vinblastine and vincristine did not exceed 0.0005% of the dry matter (Neumann et al., 2009). Earlier efforts to produce these useful compounds more efficiently relied on plant tissue culture and recombinant DNA technologies (Carew, 1966). Recently, several studies have used next-generation sequencing (NGS) tools to get a new vision on the eukaryotic transcriptome size and complexity (Wang et al., 2009). RNA-Seq has emerged as a powerful platform for providing an enormous and accurate millions of transcriptome sequence data. Researchers utilized NGS to illustrate the function of novel genes at a high-throughput basis (Oliver, 2002; Yegnasubramanian, 2013). The main advantages of RNA data analysis are the accuracy, low cost, high coverage, and great resolution compared to traditional methods (Kukurba & Montgomery, 2015; Park et al., 2016). RNA-seq analysis of gene expression have played a critical role in defining TIA encoded genes in C. roseus (El-Domyatiet al., 2017). This study was designed to detect spatial and temporal transcriptome profiling in C. roseus with emphasis on TIA pathway across different organs.

Materials and Methods

RNA-Seq data collection: Transcriptome sequences of C. roseus (Little Bright Eyes cultivar) were obtained from SRA (NCBI) with accession numbers of SRR122239 (flower), SRR122243 (immature leaf), SRR122251 (mature leaf), SRR122253 (stem), SRR122254 (root), and SRR122257 (hairy root). Data quality was measured using fastQC program after trimming adapter sequences. Read length of ≤ 50 bp and low qualities reads (≥ 20 for all bases) were removed.
Sequence assembly, genes annotation and clusters analysis: Recovered RNA-Seq reads were de novo assembled at kmer value = 50 and analyzed using Trinity RNA-Seq assembly (version r2013-2-25) and the tuxedo suite software packages. The relative abundance of the resulted transcripts was estimated using RSEM (v1.1.6). For differential expression analysis, gene clusters formed for each organ were analyzed via EdgeR software (version 3.0.0, R version 2.1.5). Blast-2-GO software (version 2.3.5) was used to detect differentially expressed (DE) transcripts. Clusters with functional enrichment were identified by setting a significant Pearson’s correlation through permutation analysis (Brown et al., 2006).

Validation of RNA-Seq datasets: Biological samples were collected from stem, hairy root, flower, root, immature and mature leaves (three months old) of C. roseus SunStorm™ cv. Apricot. Plants were grown at 25°C in growth chamber with 12 hours light as described (Brenchley et al., 2012). RNAs were isolated using Rneasy™ kit and relative expression of the SLS, sgd, STR1, OMT and DAT genes of C. roseus was detected via qRT-PCR to validate the RNA-Seq data. To synthesize cDNA, 2 μg of total RNA, oligo (dT) and SuperScript III Reverse Transcriptase (Invitrogen™ cat. no. 18080044) was used. Primer BLAST was used to design specific primers (Table 1) for qRT-PCR. Templates were amplified to recover 196 bp fragment of the C. roseus actin gene (accession no. EF688556). qRT-PCR was done according to Ramadan et al. (2019). Gene relative expression to the internal reference control was calculated using equation : \( 2^{-\Delta\Delta CT} = \frac{\text{CT target gene} - \text{CT internal control}}{\text{CT reference gene} - \text{CT internal control}} \) (Thomas & Kenneth, 2008).

Table 1. Primer sequences along with the annealing temperatures and expected band sizes (bp) to be utilized in validating RNA-Seq dataset of C. roseus via qRT-PCR. The actin gene was used as the house-keeping gene (196 bp).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Acc. no.</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Anneal temp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGD</td>
<td>EU072423</td>
<td>AGCTTTTGTAGGAAGACGTC</td>
<td>CGTAACCCGGAGATTCCGA</td>
<td>60°C</td>
</tr>
<tr>
<td>OMT</td>
<td>EF444544</td>
<td>TCTACGGCCTAATGCGTT</td>
<td>AAGTTGAACGGGTCAGGCA</td>
<td>60°C</td>
</tr>
<tr>
<td>DAT</td>
<td>AF053307</td>
<td>TGAAGATTGGCTGCTCTT</td>
<td>TTCTATGGCTTCGAGGGA</td>
<td>58°C</td>
</tr>
<tr>
<td>SLS</td>
<td>KF309242</td>
<td>AGGACAACAAAGTTAGGGCG</td>
<td>CTTGTGTCGATCGCAACT</td>
<td>58°C</td>
</tr>
<tr>
<td>STR1</td>
<td>Y10182</td>
<td>AGGCAGATGTTCTTTTGT</td>
<td>ACCCAAAATGGCCATCAGA</td>
<td>60°C</td>
</tr>
<tr>
<td>Actin</td>
<td>EF688556</td>
<td>TGGTCGTCCAAGACACACTG</td>
<td>CTCTCACGGGCAACAGGA</td>
<td>60°C</td>
</tr>
</tbody>
</table>

Table 2. Statistics analysis of Catharanthus roseus RNA-Seq data. F=flower, IML=immature leaf, ML=mature leaf, S=stem, R=root, HR=hairy root.

<table>
<thead>
<tr>
<th>Transcriptome data file (a)</th>
<th>Organ organs (b)</th>
<th>Total no. of reads (c)</th>
<th>No. of mapped reads (d)</th>
<th>% of mapped reads (e)</th>
<th>No. of unmapped reads (f)</th>
<th>% of unmapped reads (g)</th>
<th>Total no. of transcripts (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRR122239.fastq</td>
<td>Flower</td>
<td>31645190</td>
<td>28970625</td>
<td>91.55%</td>
<td>2674565</td>
<td>8.45%</td>
<td>40344</td>
</tr>
<tr>
<td>SRR122243.fastq</td>
<td>Immature leaf</td>
<td>29918941</td>
<td>27580411</td>
<td>92.18%</td>
<td>2338530</td>
<td>7.82%</td>
<td>39952</td>
</tr>
<tr>
<td>SRR122251.fastq</td>
<td>Mature leaf</td>
<td>25325632</td>
<td>2317495</td>
<td>92.07%</td>
<td>2008137</td>
<td>7.93%</td>
<td>40336</td>
</tr>
<tr>
<td>SRR122253.fastq</td>
<td>Stem</td>
<td>28903929</td>
<td>26398362</td>
<td>91.33%</td>
<td>2505567</td>
<td>8.67%</td>
<td>40936</td>
</tr>
<tr>
<td>SRR122254.fastq</td>
<td>Root</td>
<td>30896527</td>
<td>29222863</td>
<td>94.58%</td>
<td>1673664</td>
<td>5.42%</td>
<td>41281</td>
</tr>
<tr>
<td>SRR122257.fastq</td>
<td>Hairy root</td>
<td>25316070</td>
<td>23404974</td>
<td>92.45%</td>
<td>1911123</td>
<td>7.55%</td>
<td>41806</td>
</tr>
</tbody>
</table>

Clusters of gene expression in different organs: Differential expression (DE) of transcripts across C. roseus organs was detected from RNA-Seq data. To confirm the differences statistically, likelihood ratio was used to compare RPKM-derived read counts with a threshold of ≥ 2 RPKM value. Two-fold expression rate of transcripts changes and 10^3 of false discovery rate were specified. The number of differentially expressed transcripts resulted from de novo assembly was 2,266 transcripts. Pearson’s
correlation was used to analyze clusters in order to sort out expression profiles of DE transcripts and subject them to hierarchical clustering (Fig. 1). They were sorted out based on the relative abundance of transcripts across different organs. The cluster analysis of gene expression levels provided information for transcript expression patterns and pathways across different organs. In general, the number of upregulated clusters was 16 in immature leaf, while 13 in flower, 12 in mature leaf, 6 in stem, 6 in root, and 4 in hairy roots. Six out of these upregulated clusters (nos. 7, 22, 94, 97, 100 and 154; Fig. 2) contained the eight genes of TIA pathway.

Fig. 1. Hierarchical cluster analysis of gene expression across different C. roseus organs.
Fig. 2. Six selected clusters of DE transcripts from different *C. roseus* organs.
Fig. 3. Indole Alkaloid biosynthesis in TIA Pathway (El-Domyati et al., 2017).
Analysis of differentially expressed transcripts: According to TIA pathway (Fig. 3), upregulation was found in the three genes in the beginning of the TIA pathway in mature leaf, flower and root (Fig. 4). These genes encode enzymes secologanin synthase (SLS) (comp28933_c0), strictosidine synthase (STR1) (comp11842_c0), and strictosidine beta-glucosidase (SGD) (comp16844_c0) that exist in cluster 94.

In flower, the accumulation of transcripts encoding tabersonine 16-O-methyltransferase (OMT), deacetylvinodine 4-O-acetyltransferase (DAT), hydroxytabersonine N-methyltransferase (NMT) and peroxidase n1 (PRX1) seem not to support the accumulation of vindoline, vinblastine and vincristine. Therefore, it is suggested that the TIA pathway in flower is directed towards the accumulation of other alkaloids such as ajmaline, serpentine, raucuffarine or sarcapeine. However, in mature leaf, the accumulation of these transcripts was very high, indicating the possible high accumulation of vincristine, vinblastine, vindoline and possibly catharanthine. In hairy root, accumulation of previous transcripts was not observed, which indicated that vincristine, vinblastine, vindoline unlikely be accumulated in these organs. In stem, accumulation of these genes does not support accumulation of the latter alkaloids except for deacetylvinodine 4-O-acetyltransferase (DAT) (Fig. 4). In immature leaf, high accumulation of geraniol 8-hydroxylase (comp34847) was observed with no accumulation of secologanin synthase (SLS), indicating the possible accumulation of asperuloside in immature leaf (Fig. 5).

RNA-Seq data Validation: To validate RNA-Seq data, relative gene expression analysis was done via qRT-PCR. The data showed the same expression patterns of the five transcripts SLS, sgd, STR1, OMT and DAT in the RNA-Seq datasets (Fig. 6).

Discussion

The biosynthesis of _C. roseus_ terpenoid indole alkaloid has been studied for decades. Although there is tremendous progress, many steps are still not deciphered. To date, _C. roseus_ is the only natural resource for antitumor agents, vinblastine and vincristine. However, their contents in plants are very low. Furthermore, the chemical synthesis of these components is far from being applicable for commercial-scale production (Zhu et al., 2015). Many investigators aimed to detect the accumulation of indole alkaloids in different _C. roseus_ organs (Roepke et al., 2010; Zhu et al., 2015; Benyammi et al., 2016). It was also reported that some indole alkaloids like vindoline and catharanthine in a given organ can be transferred to another organ. Accumulation of these alkaloids in _C. roseus_ mainly take place in young leaves and stems, whereas synthesis of others such as ajmalicine and serpentine mainly occurs in roots (Liu et al., 2016). The main aims of this study are to understand TIA biosynthetic pathways in _C. roseus_ and detect the organ of which genes involved in this process are regulated. Transcriptome data of _C. roseus_ collected from NCBI indicates that DE transcripts in each cluster differ in their expression levels across _C. roseus_ organs. The enzymes expressed from these DE transcripts act in synthesizing TIA compounds. The six DE transcripts for TIA biosynthesis in this study were subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. There are many reports that clarified that high accumulations of vindoline, vinblastine and vincristine are mainly in mature leaf whereas catharanthine was highly accumulated in root (Malik et al., 2013; Srivastava et al., 2014). Our findings supported the previous investigation due to the expression profile of _SLS, STR1, sgd, OMT, DAT and PRX1_ genes, but we found no expression of _NMT_ gene in mature leaf although it was found in root and flower. Therefore, we suggest that the NMT protein may be transferred from root to leaf as a step towards the completion of the TIA pathway. The transferred TIA proteins were proven in many reports (Yamamoto et al., 2016). Also, our finding supports previous investigation regarding the accumulation of catharanthine in root due to the high expression of _SLS, STR_ and _sgd_ genes in root, while no or low expression of other TIA pathway genes was observed.

Another interesting finding was that _GSH_ transcript was highly accumulated in immature leaf. This outcome suggests: (1) the accumulation of asperuloside TIA product in immature leaves tissues due to the activity of _GSH_ enzyme and absence of _SLS_ transcript activity and (2) the transferring of _GSH_ enzyme from immature leaves to mature leaves tissues to open TIAs pathways in the presence of _SLS_ transcript activity.

Although we detected the possible location of expression of some _TIA_ gene in organs by transcriptome analysis, indole alkaloids accumulation in plant organs is still ambiguous as the accumulation of some indole alkaloids in a given organ is not associated with the location of gene expression or with the accumulation of the encoded enzyme in the same organ. Rather, the enzymes are likely transferred from one organ to the other in order to complete the TIA pathway in the distal organ. However, many steps of Tabronicine accumulation are still unknown.

Conclusions

Transcriptome analysis of _C. roseus_ enabled us to detect the expression levels of six genes related to TIA biosynthesis pathway across different organs. Results showed that possibility of transfer of formed enzyme in one organ and transfer to other organs to work and restore the TIA pathway. The highest expressions level of most TIA key genes took places in root, flower and mature leaf. These findings may be useful in future to determine the precise mechanism of TIA biosynthesis in _C. roseus_ organs as well as link coexpressed genes and transcription factors involved in this process. Furthermore, it is important to investigate processes facilitating transport of enzymes from one organ to the other.

Fig. 5. Expression of geraniol 8-hydroxylase (*G8H*) transcripts related to the Monoterpenoid biosynthesis in *C. roseus* organs.
Fig. 6. Relative gene expression analysis of SLS, sgd, STR1, OMT, and DAT transcripts of C. roseus in different organs.

Acknowledgments

This project was funded by the Deanship of scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. G-138-130-38. The authors, therefore, acknowledge with thanks DSR technical and financial support.

References


(Received for publication 20 November 2018)