



OPEN

De Novo transcriptome assembly and differential expression analysis of *catharanthus roseus* in response to salicylic acid

Narges Soltani¹, Farhad Nazarian Firouzabadi^{1✉}, Alireza Shafeinia², Masoud Shirali^{3,4} & Ayeh Sadat Sadr^{5✉}

The anti-cancer vinblastine and vincristine alkaloids can only be naturally found in periwinkle (*Catharanthus roseus*). Both of these alkaloids' accumulations are known to be influenced by salicylic acid (SA). The transcriptome data to reveal the induction effect (s) of SA, however, seem restricted at this time. In this study, the de novo approach of transcriptome assembly was performed on the RNA-Sequencing (RNA-Seq) data in *C. roseus*. The outcome demonstrated that SA treatment boosted the expression of all the genes in the Terpenoid Indole Alkaloids (TIAs) pathway that produces the vinblastine and vincristine alkaloids. These outcomes supported the time-course measurements of vincristine alkaloid, the end product of the TIAs pathway, and demonstrated that SA spray had a positive impact on transcription and alkaloid synthesis. Additionally, the abundance of transcription factor families including bHLH, C3H, C2H2, MYB, MYB-related, AP2/ERF, NAC, bZIP, and WRKY suggests a role for a variety of transcription families in response to the SA stimuli. Di-nucleotide and tri-nucleotide SSRs were the most prevalent SSR markers in microsatellite analyses, making up 39% and 34% of all SSR markers, respectively, out of the 77,192 total SSRs discovered.

Catharanthus roseus, a diploid ($2n = 16$) plant native to Madagascar, belongs to the Apocynaceae family. Despite the fact that it is grown as an ornamental plant in many tropical countries^{1,2}, commercial alkaloids are rich in the aerial and underground tissues of this medicinal herb³⁻⁵.

A wide range of plant driven alkaloids such as ajmalicine, serpentine, vinblastine and vincristine have been extensively used in clinical therapy^{1,6}. Regarding the plant kingdom, *C. roseus* solely synthesizes vinblastine and vincristine alkaloids which are used in cancer treatments^{7,8}. However, the synthesis rate and accumulation of both alkaloids in *C. roseus* are low and insufficient for the economic benefit of pharmaceutical industries. To this end, attempts to synthesize the aforementioned alkaloids, have not been successful yet⁸⁻¹¹. The *C. roseus* valuable alkaloids are synthesized in the Terpenoid Indole Alkaloids (TIAs) pathway. It's interesting to note that the plant's developmental stages and the cell's reaction to biotic and abiotic stimuli alter the expression of TIAs pathway genes^{7,8,12,13}. Many transcriptional activators and repressors regulate the transcription level of TIAs pathway genes^{8,14-17}. Therefore, revealing the biosynthetic pathway of TIAs and identification of the regulators of these compounds can be useful in metabolic pathway engineering^{1,8}. In recent years, several studies have been performed to identify genes and regulators of the TIAs pathway^{1,7,8,15,16,18,19}. Despite tremendous advances, there is a need to understand and identify other genes and regulators involved in the TIAs pathway⁸.

The use of Next-Generation Sequencing (NGS) technology in analyzing genomic and transcriptomic resources in all organisms, including non-model species, enables the identification of genes involved in secondary metabolism pathways^{8,20}. Recent NGS platforms have revolutionized sequencing methods to identify uncharacterized genes/proteins involved in complex pathways⁷. For instance, using the de novo method, Verma, et al.¹ investigated the transcripts of leaf, flower and root tissues in *C. roseus*, resulting in the identification of 59,220 unique

¹Production Engineering and Plant Genetics Department, Faculty of Agriculture and Natural Resources, Lorestan University, P.O. Box 465, Khorramabad, Iran. ²Department of Plant Production & Genetics, Faculty of Agriculture, Agricultural Sciences & Natural Resources, University of Khuzestan, Mollasani, Iran. ³Agri-Food and Biosciences Institute, Hillsborough BT26 6DR, UK. ⁴School of Biological Sciences, Queen's University Belfast, Belfast BT9 5AJ, UK. ⁵South of Iran Aquaculture Research Institute (SIARI), Iranian Fisheries Science Research Institute, Agricultural Research Education and Extension Organization (AREEO), Ahvaz, Iran. ✉email: nazarian.f@lu.ac.ir; a.sadr@areeo.ac.ir

Attribute	CD- Hit	Trinity
Total trinity genes	47,640	64,531
Total trinity transcripts	74,301	120,739
GC (%)	39.98	40.11
Contig N50	2201	2273
Average contig length	1479	1582
Total assembled bases	109,874,530	191,022,812

Table 1. Trinity assembly stats report of assembled contigs.

transcripts with an average length of 1284 bp. Almost 65% of the transcripts showed homology to the deposited sequences in the databases, while 35% of the unigenes were specific to the bisindole alkaloids synthesis activity in the plant leaves and roots organs, whereas enzymes associated with vindoline and vinblastine synthesis were solely found in the plant aerial tissues. Furthermore, Verma, et al.¹ identified 11,620 simple sequence repeats (SSRs) markers and 1820 transcription factor (TFs) in *C. roseus*.

Kellner, et al.⁷ generated a partial genome assembly for *C. roseus* aiming to understand the monoterpenoid indole alkaloids (MIAs) biosynthesis pathway⁷. In the past 10 years, not only the MIA pathway leading to the formation of strictosidine, but several downstream pathway branches, and previously undescribed TFs and transporters have been characterized^{21,22}.

Jasmonic acid (JA) and salicylic acid (SA) among other elicitors have been shown to alter the expression of TIAs pathway genes^{11,23,24}. It has already been reported that most alkaloid regulators in *C. roseus* are induced by JA. Nonetheless, they exert cell-specific effects, since some TFs exclusively regulate the specific part of the MIA pathway^{21,22}.

It has been documented that SA induces the expression of TIAs pathway genes and the synthesis of vindoline, vinblastine and vincristine alkaloids^{11,23,25,26}. Spraying SA on *C. roseus* leaves resulted in a significant increase in the expression of TIAs pathway genes, which in turn caused the accumulation of catharanthine, vinblastine and vindoline alkaloids^{11,27}. Also, spraying SA on leaves in *C. roseus* plants facing water²⁸ and salinity²⁵ stresses, reduced the effects of stress and increased the total amount of alkaloids, more specifically a significant increase in vincristine, and vinblastine production rate. Nevertheless, despite the beneficial effects of the application of SA, higher concentrations (> 1 mM) of SA resulted in decreasing the amount of alkaloids, suggesting that the induction role of SA seems to depend on the concentration of SA²⁹. Therefore, it can be anticipated that the expression pattern of associated TIAs pathway genes following SA treatments by RNA-Seq, will further facilitate the unravelling and identifying genes involved in the biosynthesis of complex molecules such as vincristine and vinblastine.

Results

Quantitative assessment, Sequencing and de novo assembly. In order to generate a comprehensive transcriptome of *C. roseus* leaves in response to 0.1 mM SA treatment 24-h post-treatment, cDNA libraries were constructed in three biological replicates and sequenced. In total, 191,022,812 bases were generated by RNA-Seq (Table 1). Mean unreads nucleotides (N) were reported to be zero in all samples. The average length of the genes in all the libraries was 1277 bp. Protein coding genes were assessed using the BUSCO pipeline. It was shown that the majority of the eudicot's core genes had been successfully recovered. Specifically, by identifying 2326 total BUSCO groups, 46.6% were complete single-copy, 45.1% were complete duplicated, 2% were fragmented, and 6.3% were missing.

Annotation and taxonomy of unigenes. The unigenes' annotation was done using BLAST against KAAS, Uniprot, and NCBI non-redundant (NR) protein databases. Out of 74,301 transcripts identified, 44,688 (60%), 36,292 (49%), 45,111 (61%), and 22,919 (31%) from all the libraries were annotated against KAAS, Uniprot, NCBI NR protein, and GO databases, respectively (Fig. 1). Also, the Blast results showed 63% and 64% similarity with *Arabidopsis thaliana* in BLASTP and BLASTX, respectively. From 56,939 inputs (Quarry), 47,398 (83%) unitranscripts were allocated to Eudicots based on GhostKOALA analysis³⁰. According to GhostKOALA server, genetic information processing had the highest number in the functional classification of Protein families. Sections such as secondary compounds, amino acid metabolism, and terpenoid metabolism were also identified (Fig. 2).

Gene ontology (GO) classification and Functional KEGG analysis. To determine the functional classification of unigenes, WEGO was used³¹. Results of this type of analysis led to the identification of 55 functional categories. The unigenes were divided into three categories, including cellular components, molecular function, and biological processes. In the biological process category, metabolic processes, cellular processes and biological regulation were highly represented groups. Moreover, binding and catalytic activity constituted the largest proportion of molecular function and cellular component categories, respectively. Annotation of unigenes in the KEGG database was done using KAAS, and *Arabidopsis* as the reference genome. Metabolic pathways, biosynthesis of secondary metabolites, and biosynthesis of amino acids such as tryptophan metabolism, tyrosine metabolism, and phenylalanine metabolism pathways were identified (Supplementary 2).

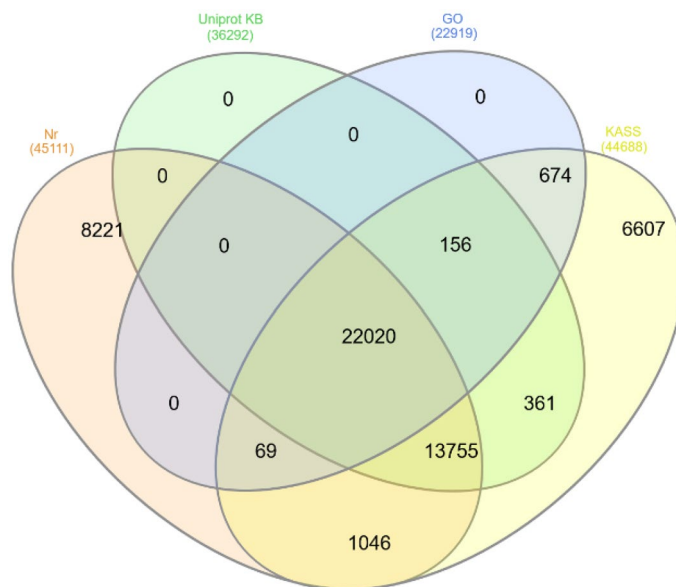


Figure 1. Venn diagram showing the number of differentially expressed transcripts homologous to known annotated genes from KASS, GO, Uniprot KB and NCBI non-redundant (NR) protein databases.

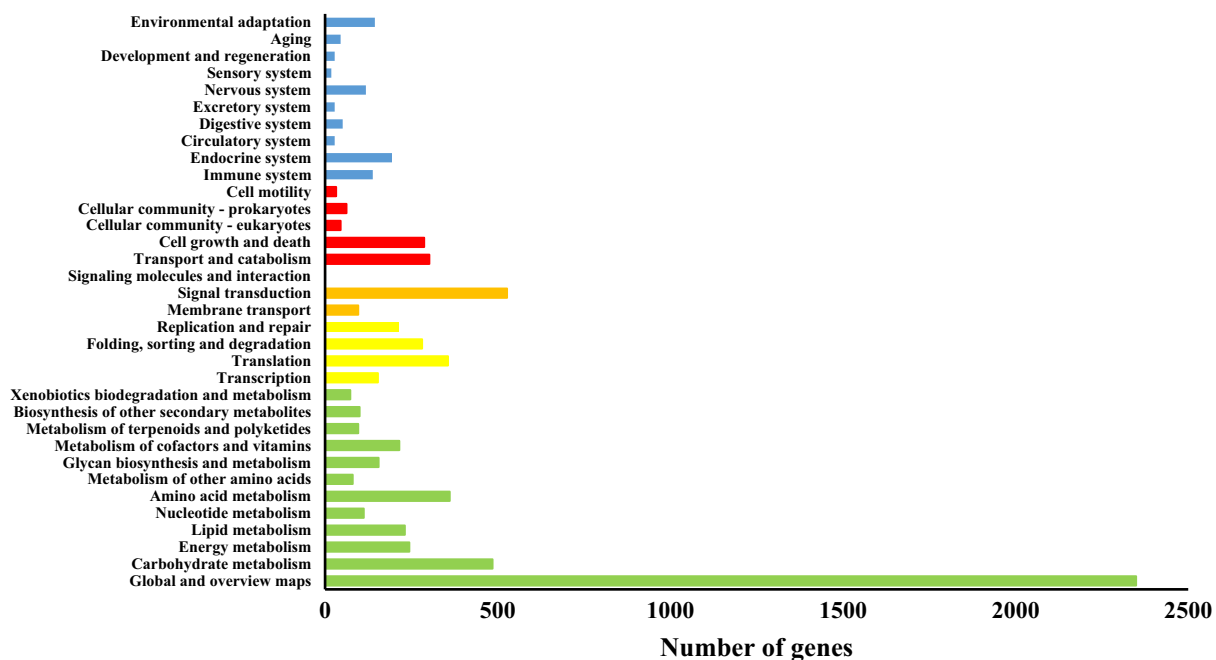


Figure 2. The result of annotation in GhostKOALA (green: Metabolism, yellow: Genetic Information Processing, orange: Environmental Information Processing, red: Cellular Processes and blue: Organismal Systems).

Identification and classification of differentially expressed unigenes. Differential expression analysis was performed using EdgeR package with $FDR < 0.05$ and $\log_2 \text{fold-change} \geq 1$ ³². In total, 8348 DEGs between the control and SA-treated libraries were identified in which 3892 DEGs were up-regulated and 4456 DEGs were down-regulated (Fig. 3). At the 95% significance threshold, unigenes with different expression levels were identified (Supplementary 3).

GO and Annotation of DEGs. Following identifying genes with different expression patterns between treatment and control samples, gene ontology was performed by goseq package³³. The results of this analysis led to the identification of 81, 5 and 52 terms for biological processes, cellular components and molecular functions, respectively (Supplementary4). Among genes associated with biological processes, the GO terms related to biological process, metabolic process and biological regulation had a significantly higher level of expression.

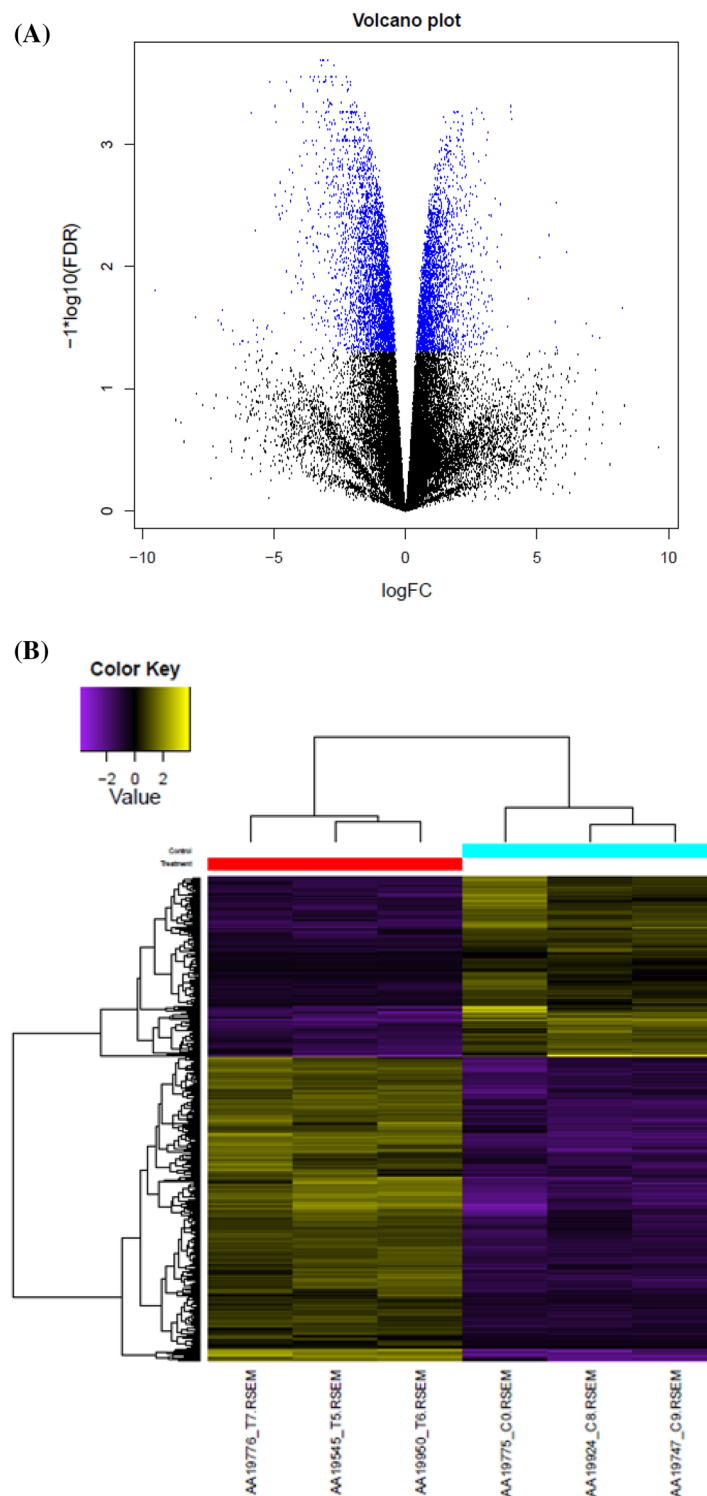


Figure 3. (A): A volcano plot illustrating the Log₂-fold change between the treatments versus control plants. Blue dots represent significantly DEGs between treatment (24-h post 0.1 mM salicylic acid) and control (distilled water treatment) plants, (B): Heat Map plot of differential expression genes between the treatment and controls.

The genes associated with metabolic pathways, biosynthesis of secondary metabolites, biosynthesis of amino acids, specifically tryptophan metabolism, tyrosine metabolism, and phenylalanine metabolism were differentially expressed. There were also differentially expressed unigenes related to the biosynthesis of terpenoid,

monoterpenoid biosynthesis, diterpenoid biosynthesis, phenylpropanoid biosynthesis, flavonoid biosynthesis, flavone and flavonol biosynthesis, indole alkaloid biosynthesis, isoquinoline alkaloid biosynthesis, and tropane, piperidine and pyridine alkaloid biosynthesis (Supplementary 5).

Genes associated with terpenoid compounds. In the pathway related to terpenoid compounds, 19 DEGs between control and treatment samples were identified, including *DXS*, *DXR*, *ISPE*, *ISPH*, *ISPF* and *GGPS*. These enzymes are at the beginning of the terpenoid pathway and are the most important enzymes in the biosynthesis pathways derived from the Methyl Erythritol Phosphate (MEP) pathway. Information and names of genes with different expressions in the biosynthetic pathway of the aromatic amino acids tryptophan, phenylalanine and tyrosine can be seen in supplementary 5. The results of this analysis showed differences in the expression of anthranilate synthase and tryptophan synthase genes. Also, the chorismate mutase gene was differentially expressed. Differences in the expression of genes related to alkaloid compounds were also noticed. The production of indole terpenoid alkaloids is specific only to *C. roseus* and is not synthesized in other plants such as *Arabidopsis*, which has been used as a template in the analysis. For this reason, the protein sequences of TIAs pathway enzymes were obtained from the UniProtKB/Swiss-Prot. The *C. roseus* database was prepared for the TIAs pathway and the sequence of transcripts was compared and analyzed. As a result of this analysis, all genes in the TIAs pathway were identified (Fig. 4).

Validation of DEGs by qRT-PCR. To confirm and validate the accuracy and reliability of the transcriptome analysis data, seven candidates unigenes associated with the TIAs pathway were chosen for qRT-PCR analysis. The results of qRT-PCR analysis for selected hub genes were previously published²⁷. However, to avoid repeated sampling errors, relative fold changes were used from our previous study. Consistent with RNA-Seq analysis, fold change expression of DEGs of the qRT-PCR analysis for selected key genes of the TIAs pathway showed an increase in expression following SA treatment plants in comparison to the control sample (Fig. 5).

Vinblastine and Vincristine content. Vinblastine and vincristine are effective agents for treating various types of cancers, including leukemias, lymphomas, and other solid tumors³⁴. To evaluate the effect of SA treatment on the production rate of these alkaloids, the amount of vinblastine and vincristine was measured at 1st, 2nd, 3rd and 7th days post-treatment by HPLC analysis (Fig. 6). It was found that SA treatment was effective in increasing the amount of vincristine, the final product of the TIAs pathway. As it can be seen from Fig. 6, SA treatment led to a significant increase in vincristine production. Although the amount of vinblastine significantly increased on day 2nd and 3rd post SA treatment, on day 7th the amount of vinblastine dropped sharply.

Identification of TFs and SSR markers. Overall, we found 1895 TFs of which 710 TFs had a differential expression, belonging to different families. bHLH, C3H, C2H2, MYB, MYB-related/AP2/ERF, NAC, bZIP, and WRKY were the most common TFs (Supplementary 6).

The *C. roseus* transcript sequences were searched for microsatellite loci using the MISA identification tool (MISA)³⁵. Overall, 77,192 SSRs were identified. Among these SSRs, the di-nucleotides and tri-nucleotide SSRs were the most abundant SSR markers, accounting for 39% and 34% of the total SSR markers, respectively (Supplementary 7). There was a low number of hexa- (5%) and penta-nucleotide (4%) SSRs.

Discussion

Omics data for medicinal plants is limited in comparison to crop plants. Transcript analysis by NGS techniques has made it possible to identify genes associated with pathways involved in the production of valuable alkaloids⁸. *C. roseus* is one of the most important medicinal plants that is considered as a model plant for secondary metabolome studies³⁶. Apart from environmental and culture conditions, elicitation is also considered to be an important factor in promoting alkaloid production in medicinal plants. For instance, the use of fungi/bacteria as elicitors for improving alkaloid production in many medicinal plants has been well documented^{11,23,37,38}. Identification and annotation of alkaloid biosynthesis in *C. roseus* are of immense importance. Gene-to-metabolite networks in *C. roseus* cells have shown that genes play a major role in metabolites accumulation¹⁶. Salicylic acid has been reported to affect the accumulation of *C. roseus* metabolites such as vindoline, vincristine, and vinblastine^{11,25}. Consequently, exogenous application of SA can increase the rate of alkaloid production in many plant species. For instance, medicinal plants subjected to SA treatments display a wide range of complex molecular responses, including changes in gene expression, modulation of transcriptional regulation, and signal transduction related to secondary metabolites pathways^{39–43}. In this study, treatment of *C. roseus* leaves with SA resulted in altering the expression of major genes involved in indole terpenoid alkaloids biosynthesis. The amount of vincristine production was also significantly increased following SA treatment, whereas vinblastine accumulation decreased dramatically on day 7th. This could be due to the vinblastine has been converted to vincristine at high capacity on day 7th^{2,6,38,44}. Currently some partial RNA-Seq data of *C. roseus* are available where one can search for key genes associated with metabolic pathways leading to the production of important alkaloids^{1,8,45,46}. Despite many efforts to identify key regulators of TIAs biosynthetic pathway, there are ambiguities regarding intermediate compounds of TIAs biosynthetic pathway^{1,8}. We performed de novo transcriptome analysis using the Illumina HiSeq 2500 platform to investigate and identify functional genes in response to SA treatment in *C. roseus*. The results showed that 83% of the transcripts assigned to the dicots. More than 60% of unigenes were annotated in databases (Fig. 1) with the highest similarity of 64% to *Arabidopsis*. Similarly, transcriptome analysis of *C. roseus* revealed 38,380 annotated unigenes with the highest similarity to *Arabidopsis* transcriptome¹. The results of comparing de novo transcript on medicinal plants database against *C. roseus* showed 94% similarity. Carbohydrates, flavonoids, saponin and alkaloids are the main constituents of *C. roseus* secondary metabolites^{2,38}. Consequently,

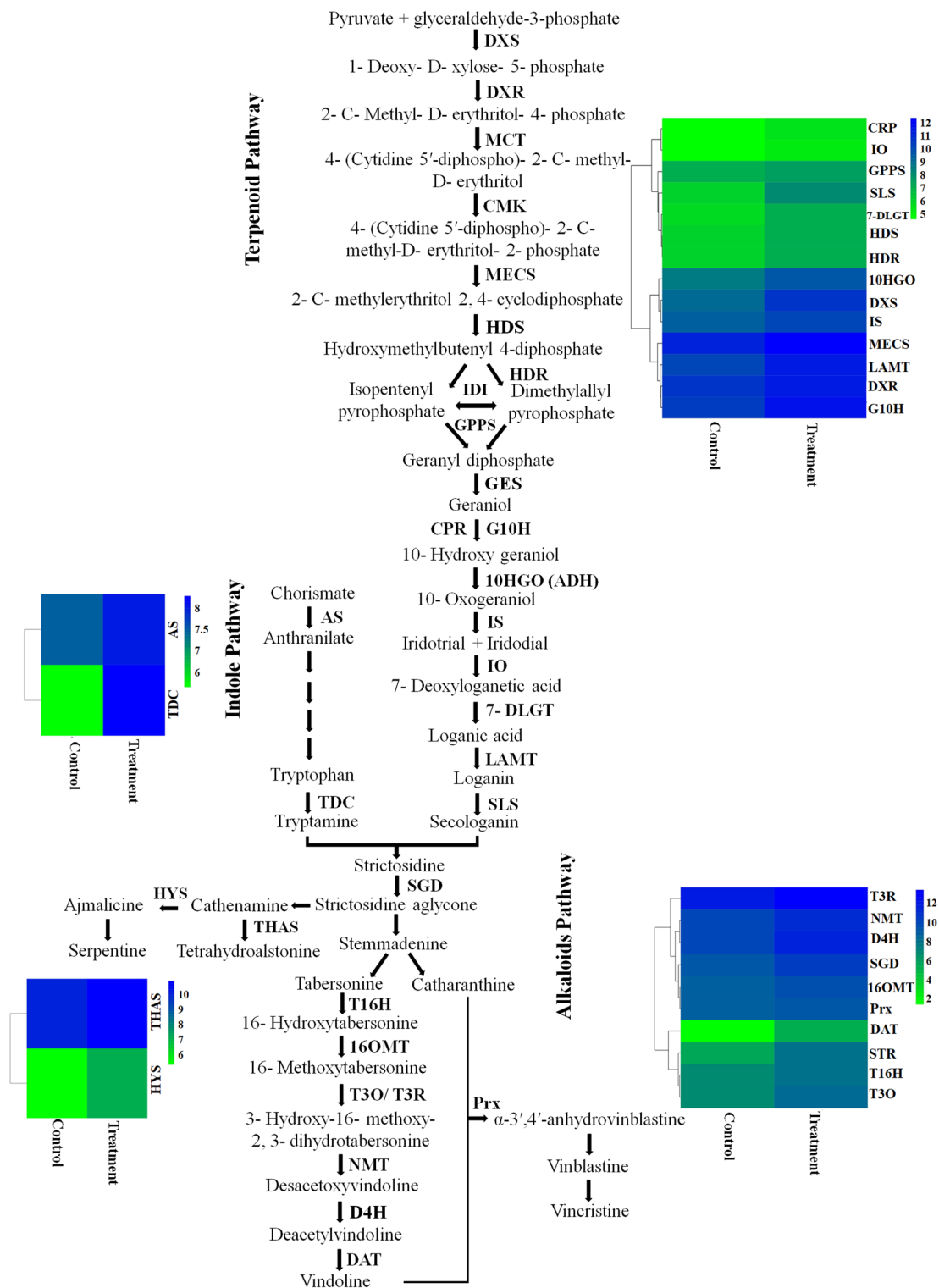


Figure 4. Expression patterns of *C. roseus* genes involved in the terpenoid indole alkaloids biosynthetic pathways.

the highest percentage of annotated genes belonged to biological functions. Similarly, Verma, et al.¹ and Pan, et al.⁸ identified the same pathways following treatment of *C. roseus* plants with methyl jasmonate. *C. roseus* terpenoid indole alkaloids are derived from the MEP and shikimate pathways^{2,44}. The product of the terpenoid and indole pathways enters the synthesis pathway of terpenoid indole alkaloids with the activity of the STR enzyme. Results of this study showed that the expression of genes involved in alkaloids biosynthesis in *C. roseus* changed following salicylic acid treatment. qRT-PCR analysis showed that SA treatment increased the expression of genes

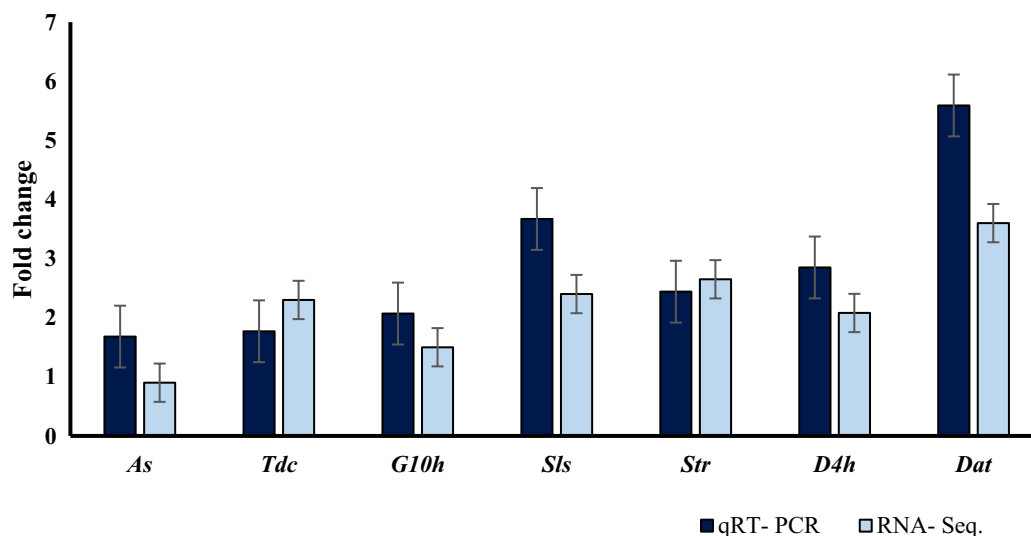


Figure 5. Comparison of qRT-PCR²⁷ and RNA-Seq of *Tdc*, *G10h*, *Sls*, *Str*, *D4h* and *Dat* genes in 24-h treatment.

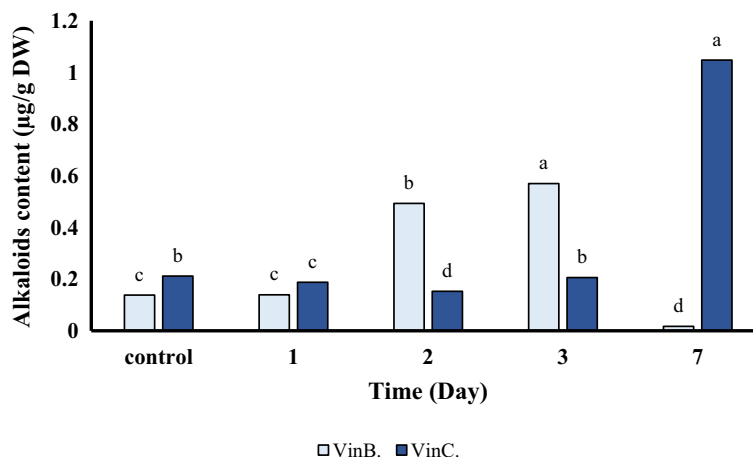


Figure 6. Effect of spraying of salicylic acid on vinblastine and vincristine content of *C. roseus*.

associated with the synthesis of terpenoid indole alkaloids (Fig. 5). Consequently, up-regulation of these genes led to vincristine and to a lesser extent vinblastine production (Fig. 6).

It was found that many of up-regulated genes in *C. roseus* in response to SA treatment, were key receptors/signaling genes or were associated with SA-mediated responses. The results of this study have confirmed the results of previous studies by indicating the positive effect of SA treatment on genes associated with TIAs pathway and its end product, vincristine. As far as the literature is concerned, until now the comparison between salicylic acid and MeJA's influence on the expression of genes involved in alkaloids biosynthesis in *C. roseus* has not been reported. However, separate pieces of research have been conducted to assess the effect of MeJA and JA on the expression of genes associated with TIAs pathway. For instance, Wei (2010)⁴⁷ showed that MeJA treatments induced the expression patterns of TIAs pathway genes in *Catharanthus roseus* seedlings. Even though MeJA and JA seem to share similar effects on the overall expression of TIAs genes, according to Rischer et al. (2006), using *C. roseus* cell suspension cultures showed that JA induced the expression of some TIAs biosynthetic genes, such as *G10H*, *SGD*, and *TDC*, but the expression of genes encoding *MAT* and *DAT* did not change. Taken together and in comparison to our results, MeJA, JA and SA treatments seem to target similar TIAs genes expression. However, further analysis is needed to unveil the effects of all three organic compounds in a single study at the same time.

TFs play an important role by modulating the expression level of their target genes. NGS methods with the aim of bioinformatics analysis have made it possible to identify both TFs and target proteins in different species^{6,48}. In the present study, 1895 TFs were identified that 710 transcription factors had differential expression. bHLH, C3H, C2H2, MYB, MYB-related, AP2/ERF, NAC, bZIP and WRKY families had the highest number of TF members in *C. roseus*⁶. Interestingly, Verma, et al.¹ identified bHLH, MYB, AP2/ ERF and WRKY TFs in a *C. roseus* transcriptome study. Also, the study of the effect of methyl jasmonate and ethylene on the synthesis

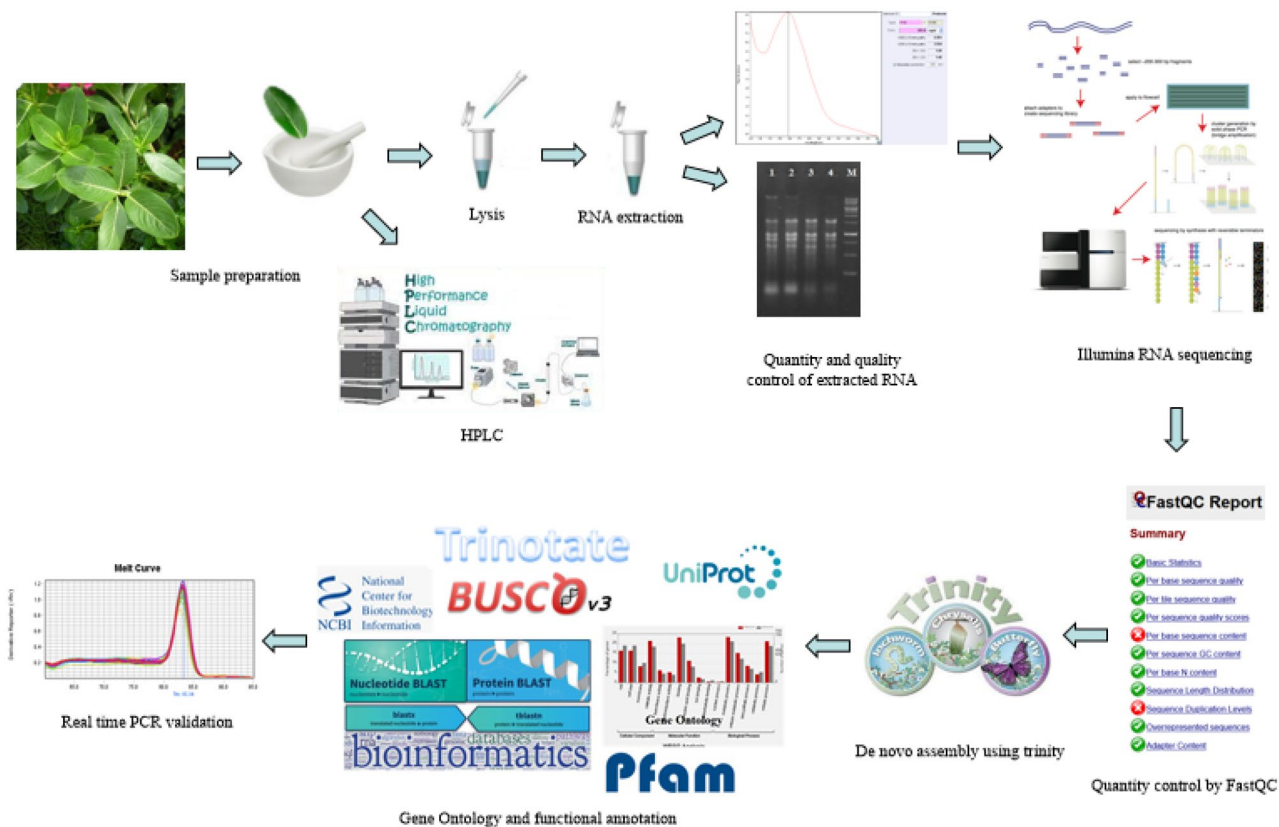


Figure 7. Schematic illustration of different steps taken for analysis.

of TIAs led to the identification of WRKY and ABC TFs⁸. It is well documented that the synthesis of terpenoid indole alkaloids is modulated by TFs related to the structural genes of the TIAs pathway^{6,49,50}. For example, two TFs, ORCA2 and ORCA3, as members of the AP2/ ERF transcription factor family, are crucial regulators TIAs Pathway^{6,51}. The overexpression of ORCA2 significantly increased the expression of many genes in the TIAs pathway resulting in a higher accumulation of catharanthine, ajmalicine, serpentine, tabersonine⁵² and vindoline⁵³. In addition to the role of TFs belonging to the AP2/ ERF family, it has been shown that CrMYB2 overexpression in the hairy roots of *C. roseus* alters the expression of ORCAs and also increases the amount of catharanthine and tabersonine⁵⁴. In fact, the bHLH transcription factor regulates the expression of structural genes, whereas ORCA3 does not regulate structural genes and may play redundant roles²². WRKY1 transcription factor binds to the Wbox element in the promoter of the *Tdc* gene, regulating the expression level of *As*, *Dxs*, *Sls*, *Sgp* and *Tdc* genes, consequently, and modulating the production of serpentine and decreased catharanthine. Therefore, WRKY is a good candidate for regulating metabolites such as serpentine and ajmalicine production^{6,14}.

SSR markers are co-dominant species-specific markers which are used in breeding programs to study genetic diversity, and kinship relationships germplasm analysis^{55,56}. Out of 77,192 SSR markers identified by MISA, three and two nucleotide repeats had the highest number of SSRs with 34% and 39%, respectively. The presented results were consistent with the results of Verma, et al.¹ where the highest percentage of SSRs belonged to two and three nucleotide repeats.

Conclusion

In addition to the importance of the *C. roseus* as the only source of vinblastine sulfate and vincristine sulfate drugs, it is also considered as a model plant in studying secondary metabolites production. This is the first study dealing with the expression analysis of TIAs pathway genes following SA treatment. The expression of all genes in the TIAs pathway producing anti-cancer alkaloids vinblastine and vincristine increased after SA treatment. These results were consistent with the measurements of vincristine alkaloid (the final product of the TIAs pathway) over time and showed that SA spray had a positive effect on both transcription and production of alkaloids. The result of this study can be used as a way to increase the anti-cancer alkaloids in *C. roseus*.

Materials and methods

The pipeline of study is shown in Fig. 7.

Plant material, growth conditions, and treatment. Periwinkle (*Catharanthus roseus* (L.) G. Don 'Pacifica XP Burgundy Halo Vinca' seeds were purchased from Pan American Seed Company (USA, SKU: 506,521). The *C. roseus* seeds were planted in pots (10 cm diameter × 10 cm height) and kept in a growth cham-



Figure 8. Seedlings of *C. roseus*.

ber (25 °C, 16 h photoperiod, and light intensity of 70% $\mu\text{mol m}^{-2} \text{s}^{-1}$). At the beginning of flowering stage (Fig. 8) as recommended^{2,57}, SA has sprayed on the leaves at concentrations of 0.01 and 0.1 mM²³. Leaf samples were collected from the third and fourth leaves according to Roepke, et al.⁵⁸ at 12, 18, 24, and 48 h post-treatment (3 replicates each). The collection of plant material, complies with relevant institutional, national, and international guidelines and legislation. Plant materials were then immediately frozen in the liquid nitrogen and stored at -80 °C for RNA extraction. Based on the results of qRT-PCR analysis from our previous study²⁷ (Supplementary 1), 24-h post SA treatment (SA: 0.1 mM salicylic acid) and control plants (C: distilled water treatment) were selected for RNA-Seq.

RNA preparation, cDNA library construction and sequencing. Total RNA was extracted from leaves using the Monarch® Total RNA Miniprep Kit (NEB Company, UK. Cat# T2010) according to the manufacturer's instructions. The extracted RNA samples were treated with DNase I enzyme to remove possible DNA contaminations. The quantity and quality of RNA samples were analyzed by using the NanoDrop 2000c Spectrophotometer (Thermo Scientific NanoDrop 2000, USA) and 1% agarose gel electrophoresis, respectively. Next, high-quality RNA samples (three biological replicates), were sent to the sequencing company (Beijing Genomes Institute, China). Furthermore, quality controls and the RNA integrity number (RIN) were assessed for all RNA samples. cDNA preparation and sequencing were performed with the use of standard Illumina kits according to the manufacturer's instructions (Beijing Genomes Institute, China). Finally, paired-end sequencing was performed by Illumina HiSeq 2500 platform with a read length of 150 nucleotide.

Quality control, processing and de novo assembly. Sequencing data quality assessment was performed by using FastQC 0.11.20⁵⁹. Trimmomatic 0.36⁶⁰ was employed to remove adapters, nucleotides and low quality readings as well as to check the quality of the nucleotides reads. De novo assembly was conducted using Trinity 2.11.0⁶¹ by default parameters, k-mer length of 32 and minimum transcripts length of 300 bp. Trinity transcripts were clustered using CD-HIT by identity cutoff of 95%⁶². Transcriptome completeness was assessed using the bioinformatics tool BUSCO v3 (Benchmarking Universal Single-Copy Orthologs) to obtain the percentage of single-copy orthologues represented in three datasets edicots_odb10⁶³. After integrating the readings, to identify the candidate Coding Sequences (CDS), all sequences were translated into protein sequences by TransDecoder tool (<http://transdecoder.github.io>). To find homology for identified transcripts and translated peptides, BLASTX and BLASTP were performed, respectively.

Functional annotation of *C. roseus* assembled unigenes. Using a Trinotate annotation pipeline (<http://trinotate.github.io/>), functional annotation of the *C. roseus* transcriptome was done. Next, the assembled unigenes of *C. roseus* were blasted against NR proteins (<http://www.ncbi.nlm.nih.gov/>) and UniProtKB/Swiss-Prot (<https://www.uniprot.org/>) databases. Then, HMMER software was used to predict protein domains against the Pfam database. Also, signal peptide and transmembrane domains were identified using the SignalP5.0 server. Biological pathway information (Available in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database) using GhostKOALA (KEGG Orthology And Links Annotation)³⁰ (<https://www.kegg.jp/blastkoala/>) and KEGG Automatic Annotation Server (KAAS) (<https://www.genome.jp/kegg/kaas/>) servers, gene ontology terms and functional classifications (biological, cellular, and molecular) related to identified homologous genes were extracted from the WEGO (Web Gene Ontology Annotation Plot) Web tool (<http://wego.genomics.org.cn/>).

Differential gene expression analysis. Initially, all high-quality readings were aligned with Bowtie2 software against the assembled transcript. Then, the frequency of integrated copies was estimated using RSEM⁶⁴. EdgeR package was used to normalize and obtain the differences in gene expression. Next, using the goseq package³³ in R software 4.1.0⁶⁵, ontological analysis and classification of terms into cellular, molecular and biological categories for genes with different expressions was performed. Finally, the KASS database was used to annotate unigenes with different expressions.

Differential gene expression related to terpenoid indole alkaloids (TIAs) pathway. Protein sequences related to the *C. roseus* TIAs pathway (available in the UniProtKB/Swiss-Prot database) were obtained. Differentially expressed transcripts were blasted against the *C. roseus* protein database to re-evaluate the expression changes of the genes in the target pathway. Finally, the diagram of the genes involved in the TIAs pathway was plotted.

qRT-PCR validation. For qRT-PCR validation, Total RNA extraction and cDNA synthetize were carried out. The full details of the qRT-PCR analysis and the results were published in Soltani et al.²⁷.

Determination of alkaloid contents. The samples were prepared according to Singh, et al.⁶⁶ with slight modifications. Briefly, leaf samples were dried in the shade and pulverized leaf tissues were three times suspended in 80% methanol solution. After drying, 10 ml of water was added to the samples and the pH was adjusted to 3.5. The aqueous phase was removed by washing three times with chloroform, the pH was raised to 8.5 by adding ammonia and after three washes with chloroform, the chloroform phase was removed. Methanol was added to the dried sample and stored in the refrigerator for further analysis.

HPLC analysis. High-performance liquid chromatography was performed using Eurospher II 100–5 C18 column with precolumn (Column 250 * 4.6 mm) and UV detector (model K-2600) with an HPLC device (KNAUER WellChrom model) at 254 nm. The mobile phase consisted of methanol: diammonium dihydrogen phosphate (5 mM) with a composition of 71:29 (pH=7.3) and a flow rate of 1 mL/min⁶⁷.

Statistical analysis. One-way analysis of variance (ANOVA) was done by using Agricolae package 1.3–5⁶⁸ in R software 4.1.0⁶⁵ to statistically compare changes of vinblastine and vincristine alkaloids content at the different sampling points. The Least Significant Difference (LSD) test was employed to compare treatment means at a confidence of 95%.

Identification of TFs and SSRs. For the identification of *C. roseus* TFs families, the TF database Plant TFDB 5.0 was used⁶⁹. The MISA web (<https://webblast.ipk-gatersleben.de/misa/>)³⁵ was used to identify the *C. roseus* microsatellites.

Data availability

All the Illumina sequencing reads have been deposited in the National Genomics Data Centre with the accession code (BioProject ID: PRJNA830709). <https://dataview.ncbi.nlm.nih.gov/object/PRJNA830709?reviewer=2nim3dscqae7pc7vrvkvj8gens>.

Received: 4 April 2022; Accepted: 12 September 2022

Published online: 24 October 2022

References

- Verma, M., Ghangal, R., Sharma, R., Sinha, A. K. & Jain, M. Transcriptome analysis of *Catharanthus roseus* for gene discovery and expression profiling. *PLoS ONE* **9**, e103583 (2014).
- Nejat, N. et al. Ornamental exterior versus therapeutic interior of Madagascar periwinkle (*Catharanthus roseus*): the two faces of a versatile herb. *Sci. World J.* **2015**(1), 19 (2015).
- Schluttenhofer, C., Pattanaik, S., Patra, B. & Yuan, L. Analyses of *Catharanthus roseus* and *Arabidopsis thaliana* WRKY transcription factors reveal involvement in jasmonate signaling. *BMC Genomics* **15**, 502 (2014).
- Sun, J. *Metabolic engineering and elucidation of the terpenoid indole alkaloid pathway in Catharanthus roseus hairy roots* (Colorado State University, 2016).
- Lawal, O. A., Ogunwande, I. A., Ibirogba, A. E., Layode, O. M. & Opoku, A. R. Chemical constituents of essential oils from *Catharanthus roseus* (L.) G. Don grown in Nigeria. *Journal of Essential Oil Bearing Plants* **18**(57), 63 (2015).
- Liu, J., Cai, J., Wang, R. & Yang, S. Transcriptional regulation and transport of terpenoid indole alkaloid in *Catharanthus roseus*: exploration of new research directions. *Int. J. Mol. Sci.* **18**, 53 (2017).
- Kellner, F. et al. Genome-guided investigation of plant natural product biosynthesis. *Plant J.* **82**, 680–692 (2015).
- Pan, Y.-J. et al. Transcriptomics comparison reveals the diversity of ethylene and methyl-jasmonate in roles of TIA metabolism in *Catharanthus roseus*. *BMC Genomics* **19**, 1–14 (2018).
- Carquejreiro, I., Noronha, H., Duarte, P., Gerós, H. & Sottomayor, M. Vacuolar transport of the medicinal alkaloids from *Catharanthus roseus* is mediated by a proton-driven antiport. *Plant Physiol.* **162**, 1486–1496 (2013).
- Wang, Q. et al. Development of efficient *Catharanthus roseus* regeneration and transformation system using *Agrobacterium tumefaciens* and hypocotyls as explants. *BMC Biotechnol.* **12**, 34 (2012).
- Pan, Q. et al. Effect of plant growth regulators on the biosynthesis of vinblastine, vindoline and catharanthine in *Catharanthus roseus*. *Plant Growth Regul.* **60**, 133–141 (2010).
- Higashi, Y. & Saito, K. Network analysis for gene discovery in plant-specialized metabolism. *Plant Cell Environ.* **36**, 1597–1606 (2013).
- Geerlings, A., Ibañez, M.M.-L., Memelink, J., van der Heijden, R. & Verpoorte, R. Molecular cloning and analysis of strictosidine β -D-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *J. Biol. Chem.* **275**, 3051–3056 (2000).

14. Suttipanta, N. *et al.* The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiol.* **157**, 2081–2093 (2011).
15. Yu, F. & De Luca, V. ATP-binding cassette transporter controls leaf surface secretion of anticancer drug components in *Catharanthus roseus*. *Proc. Natl. Acad. Sci.* **110**, 15830–15835 (2013).
16. Rischer, H. *et al.* Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. *Proc. Natl. Acad. Sci.* **103**, 5614–5619 (2006).
17. Peebles, C. A. *et al.* The expression of 1-deoxy-D-xylulose synthase and geraniol-10-hydroxylase or anthranilate synthase increases terpenoid indole alkaloid accumulation in *Catharanthus roseus* hairy roots. *Metab. Eng.* **13**, 234–240 (2011).
18. Liu, L.-Y.D. *et al.* High-throughput transcriptome analysis of the leafy flower transition of *Catharanthus roseus* induced by peanut witches'-broom phytoplasma infection. *Plant Cell Physiol.* **55**, 942–957 (2014).
19. Zhu, X., Zeng, X., Sun, C. & Chen, S. Biosynthetic pathway of terpenoid indole alkaloids in *Catharanthus roseus*. *Front. Med.* **8**, 285–293 (2014).
20. Yan, Y., Wang, Z., Tian, W., Dong, Z. & Spencer, D. F. Generation and analysis of expressed sequence tags from the medicinal plant *Salvia miltiorrhiza*. *Sci. China Life Sci.* **53**, 273–285 (2010).
21. Van Moerkercke, A. *et al.* The basic helix-loop-helix transcription factor BIS 2 is essential for monoterpenoid indole alkaloid production in the medicinal plant *Catharanthus roseus*. *Plant J.* **88**, 3–12 (2016).
22. Van Moerkercke, A. *et al.* The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Proc. Natl. Acad. Sci.* **112**, 8130–8135 (2015).
23. Kaur, J., Singh, A., Pathak, T. & Kumar, K. *Catharanthus roseus* 309–319 (Springer, 2017).
24. Zhu, J., Wang, M., Wen, W. & Yu, R. Biosynthesis and regulation of terpenoid indole alkaloids in *Catharanthus roseus*. *Pharmacogn. Rev.* **9**, 24 (2015).
25. Idrees, M., Naeem, M., Aftab, T. & Khan, M. M. A. Salicylic acid mitigates salinity stress by improving antioxidant defence system and enhances vincristine and vinblastine alkaloids production in periwinkle [*Catharanthus roseus* (L.) G. Don]. *Acta Physiol. Plant.* **33**, 987–999 (2011).
26. Mokhaberi, A., Ahmadi, J. & Mafakheri, S. The expression profile of D4H and DAT genes in *Catharanthus roseus* in response to drought, salinity and salicylic acid. *Iran. J. Genet. Plant breed.* **2**, 38–46 (2013).
27. Soltani, N., Nazarian-Firouzabadi, F., Shafeinia, A., Sadr, A. S. & Shirali, M. The expression of Terpenoid Indole Alkaloid (TIAs) pathway genes in *Catharanthus roseus* in response to salicylic acid treatment. *Mol. Biol. Rep.* **47**, 1–8 (2020).
28. Idrees, M. *et al.* Modulation of defence responses by improving photosynthetic activity, antioxidative metabolism, and vincristine and vinblastine accumulation in *Catharanthus roseus* (L.) G. Don through salicylic acid under water stress. *Russ. Agric. Sci.* **37**(474), 482 (2011).
29. Khandan-Mirkohi, A., Khalili Halbi, M., Salami, S. A. & Lesani, H. Improving effects of mild cold stress and salicylic acid on growth and physiology of periwinkle (*Catharanthus roseus* Don.). *Int. J. Hortic. Sci. Technol.* **4**, 67–78 (2017).
30. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
31. Ye, J. *et al.* WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, W293–W297 (2006).
32. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
33. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, 1–12 (2010).
34. Herradón, E., González, C., González, A., Uranga, J. A. & López-Miranda, V. Cardiovascular toxicity induced by chronic vincristine treatment. *Front. Pharmacol.* <https://doi.org/10.3389/fphar.2021.692970> (2021).
35. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585 (2017).
36. Pandey-Rai, S. *et al.* Volatile components of leaves and flowers of periwinkle *Catharanthus roseus* (L.) G. Don from New Delhi. *Flavour Fragr. J.* **21**, 427–430 (2006).
37. Pandey, S. S. *et al.* Fungal endophytes of *Catharanthus roseus* enhance vindoline content by modulating structural and regulatory genes related to terpenoid indole alkaloid biosynthesis. *Sci. Rep.* **6**, 1–14 (2016).
38. Alam, M. M., Naeem, M., Khan, M. M. A. & Uddin, M. *Catharanthus roseus* 277–307 (Springer, 2017).
39. Hadzadeh, M., Ofoghi, H., Kianirad, M. & Amidi, Z. Elicitation of pharmaceutical alkaloids biosynthesis by salicylic acid in marine microalgae *Arthrospira platensis*. *Algal Res.* **42**, 101597 (2019).
40. Kang, S.-M. *et al.* Effects of methyl jasmonate and salicylic acid on the production of bilobalide and ginkgolides in cell cultures of *Ginkgo biloba*. *In Vitro Cell. Dev. Biol.* **42**, 44–49 (2006).
41. Malarz, J., Stojakowska, A. & Kisiel, W. Effect of methyl jasmonate and salicylic acid on sesquiterpene lactone accumulation in hairy roots of *Cichorium intybus*. *Acta Physiol. Plant.* **29**, 127–132 (2007).
42. Norozi, A., Hosseini, B., Jafari, M., Farjaminezhad, M. & Palazon, J. Enhanced h6h transcript level, antioxidant activity and tropane alkaloid production in *Hyoscyamus reticulatus* L. hairy roots elicited by acetylsalicylic acid. *Plant Biosyst. International J. Deal. Asp. Plant Biol.* **153**, 360–366 (2019).
43. Yu, Z.-Z., Fu, C.-X., Han, Y.-S., Li, Y.-X. & Zhao, D.-X. Salicylic acid enhances jaceosidin and syringin production in cell cultures of *Saussurea medusa*. *Biotech. Lett.* **28**, 1027–1031 (2006).
44. Sun, J., Manmathan, H., Sun, C. & Peebles, C. A. Examining the transcriptional response of overexpressing anthranilate synthase in the hairy roots of an important medicinal plant *Catharanthus roseus* by RNA-seq. *BMC Plant Biol.* **16**, 1–14 (2016).
45. Van Moerkercke, A. *et al.* CathaCyc, a metabolic pathway database built from *Catharanthus roseus* RNA-Seq data. *Plant Cell Physiol.* **54**, 673–685 (2013).
46. Liu, J. *et al.* The combined effects of ethylene and MeJA on metabolic profiling of phenolic compounds in *Catharanthus roseus* revealed by metabolomics analysis. *Front. Physiol.* **7**, 217 (2016).
47. Wei, S. Methyl jasmonic acid induced expression pattern of terpenoid indole alkaloid pathway genes in *Catharanthus roseus* seedlings. *Plant Growth Regul.* **61**, 243–251 (2010).
48. Kodama, M. *et al.* Identification of transcription factor genes involved in anthocyanin biosynthesis in carrot (*Daucus carota* L.) using RNA-Seq. *BMC Genom.* **19**, 1–13 (2018).
49. Mertens, J. *et al.* The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiol.* **170**, 194–210 (2016).
50. Zhao, J. & Verpoorte, R. Manipulating indole alkaloid production by *Catharanthus roseus* cell cultures in bioreactors: from biochemical processing to metabolic engineering. *Phytochem. Rev.* **6**, 435–457 (2007).
51. Menke, F. L., Champion, A., Kijne, J. W. & Memelink, J. A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene *Str* interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. *EMBO J.* **18**, 4455–4463 (1999).
52. Li, C. Y. *et al.* The ORCA2 transcription factor plays a key role in regulation of the terpenoid indole alkaloid pathway. *BMC Plant Biol.* **13**, 1–17 (2013).
53. Liu, D.-H. *et al.* Enhanced accumulation of catharanthine and vindoline in *Catharanthus roseus* hairy roots by overexpression of transcriptional factor ORCA2. *Afr. J. Biotech.* **10**, 3260–3268 (2011).

54. Zhang, H. *et al.* The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in *Catharanthus roseus*. *Plant J.* **67**, 61–71 (2011).
55. Kumar, S. & Bhatia, S. A polymorphic (GA/CT) n-SSR influences promoter activity of Tryptophan decarboxylase gene in *Catharanthus roseus* L. Don. *Sci. Rep.* **6**, 1–12 (2016).
56. Shokeen, B., Choudhary, S., Sethy, N. K. & Bhatia, S. Development of SSR and gene-targeted markers for construction of a framework linkage map of *Catharanthus roseus*. *Ann. Bot.* **108**, 321–336 (2011).
57. Idrees, M., Naeem, M. & Khan, M. M. A. The superiority of cv 'rosea' over cv 'alba' of periwinkle (*Catharanthus roseus* L.) in alkaloid production and other physiological attributes. *Turk. J. Biol.* **34**, 81–88 (2010).
58. Roepke, J. *et al.* Vinca drug components accumulate exclusively in leaf exudates of Madagascar periwinkle. *Proc. Natl. Acad. Sci.* **107**, 15287–15292 (2010).
59. Andrews, S. FastQC: A quality control tool for high throughput sequence data. **17**, 2018. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (Accessed May 2010).
60. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
61. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
62. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
63. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
64. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 1–16 (2011).
65. De Mendiburu, F. & Yaseen, M. *Agricolae*: Statistical procedures for agricultural research. *R package* version 1.4.0. <https://github.com/myaseen208/agricolae> (2020).
66. Singh, D., Maithy, A., Verma, R., Gupta, M. & Kumar, S. Simultaneous determination of *Catharanthus* alkaloids using reversed phase high performance liquid chromatography. *Liq. Chromatogr. Relat. Technol.* **23**, 601–607 (2000).
67. Hisiger, S. & Jolicœur, M. Analysis of *Catharanthus roseus* alkaloids by HPLC. *Phytochem. Rev.* **6**, 207–234 (2007).
68. de Mendiburu, F. *agricolae* tutorial (Version 1.3–5). **88** (2021).
69. Jin, J., Zhang, H., Kong, L., Gao, G. & Luo, J. Plant TFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* **42**, D1182–D1187 (2014).

Author contributions

N.S. and F.N.F. contributed to the design of the work. A.S.S., N.S. and M.S. contributed to data analysis and interpretation. N.S., F.N.F. and A.S.S. contributed to drafting and editing the manuscript. A.S. and M.S. contributed to critical revision of the manuscript and approving the final version of the manuscript. All authors also participated in the finalization of the manuscript and approved the final draft. The formal identification of the plant material used by F.N.F. and A.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-20314-4>.

Correspondence and requests for materials should be addressed to F.N.F. or A.S.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022