De novo transcriptome analysis of Justicia adhatoda reveals candidate genes involved in major biosynthetic pathway

Deepthi Padmanabhan
SRMIST: SRM Institute of Science and Technology
Adil Lateef
SRMIST: SRM Institute of Science and Technology
Senthilkumar Palanisamy (mpsenthilkumar@gmail.com)
SRM University https://orcid.org/0000-0003-0987-4946
Purushothaman Natarajan
West Virginia State University

Research Article

Keywords: RNA sequencing, De novo assembly, annotation, SALMON, SSR, qRT-PCR

Posted Date: January 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1249838/v1

Abstract

Background

Justicia adhatoda is an important medicinal plant traditionally used in the Indian system of medicine and the absence of molecular-level studies on this plant hinders its wide use, hence the study was aimed to analyse the genes involved in various pathways.

Methods and Results

The RNA isolated was subjected to Illumina sequencing. *De novo* assembly was performed using TRINITY software which produced 171.064 transcripts with 55,528 genes and N50 value of 2065bp, followed by annotation of unigenes against NCBI, KEGG and Gene Ontology with 105572 annotated unigenes and 40288 non-annotated unigenes. A total of 5980 unigenes were annotated to 144 biochemical pathways, including the metabolism and biosynthesis pathways. The pathway analysis revealed the major transcripts involved in the tryptophan biosynthesis with TPM values of 6.0903, 33.6854, 11.527, 1.6959, and 8.1662 for Anthranilate synthase alpha, Anthranilate synthase beta, Arogenate/Prephenate dehydratase, Chorismate synthase and Chorismate mutase, respectively. The qRT-PCR validation of the key enzymes showed up-regulation in mid mature leaf when compared to root and young leaf tissue. A total of 16,154 SSRs were identified from the leaf transcriptome of *J. adhatoda,*which could be helpful in molecular breeding.

Conclusion

The study aimed at identifying uni-transcripts involved in the tryptophan biosynthesis pathway for its medicinal properties, as it acts as a precursor to the acridone alkaloid biosynthesis with major key enzymes and their validation. This is the first study that reports transcriptome assembly and annotation of the *J. adhatoda* plant.

Introduction

Justicia adhatoda Linn. is a perennial shrub found in the tropical regions of southeast Asia belonging to the *Acanthaceae* family. It is widely branched with pink, purple, or white flowers and can grow up to a height of 2.5mm. The plant has been widely used in the Ayurvedic and Unani systems of Indian medicine for treating various diseases such as bronchitis, common cold, asthma, cough, and tuberculosis [2]. *J. adhatoda* has multiple pharmacological properties that have been used to treat blood disorders, jaundice, mouth ulcers, vomiting, fever, and heart-related problems [31,32]. Phytochemical screening of *J. adhatoda* methanolic leaf extract revealed the presence of alkaloids, anthraquinones, cardiac glycosides, flavonoids, saponins, polyphenols, terpenoids, and phytosterols [15]. The leaf extract of *J.adhatoda* acts as a potent anti-diabetic and anti-lipidaemic agent [3] and is also found to possess various pharmacological properties such as anti-microbial, hepatoprotective, wound healing, anti-ulcer, and anti-inflammatory activity [4]. The major alkaloids present in the leaf extracts are quinazoline alkaloids such asvasicoline, adhatodine, vasicolinone, adhavasinone, and vasicol [4]. GC-MS analysisof methanolic leaf extract of *J.adhatoda* revealed the presence of various bioactive compounds such as 9, 12, 15-Octadecatrienoic acid (Z, Z, Z), n-Hexadecaonoic acid, Phytol, and Amrinone phytocompounds [15]. Anisotine and vasicoline are potent inhibitors of the main protease, which is a key component for cleaving the viral polyprotein that is targeted for treating COVID-19 [17]. The development of next-generation sequencing has been useful to elucidate the secondary metabolite and its candidate genes with lowlabor and cost-effectiveness [6]. *De novo* sequencing plays a vital role in gene discovery where a reference genome sequence is not available in a non-model organism, as it is a cost-effective method [16]. An assembly of short-read sequence data is used to identify the candidate genes involved in biosynthetic pathways,

GC content,and gene expression analysis [16]. Quantitative real-time PCR elucidates the level of gene expression that encodes the various components involved in the biosynthetic pathways. Despite the well-established role of *J.adhatoda* in the medicinal system, the genomics of this plant is the least explore [1]. Thus, the present study provides a better understanding of the biosynthetic pathways and the candidate genes that are attributed to the medicinal properties.

Materials And Methods

Plant material preparation and RNA isolation

Mature and healthy leaves were collected, and RNA extraction was carried out using TRIzol® Reagent (Invitrogen, USA). The total RNA extracted was treated with DNase A and purified with the RNeasyMinElute clean-up kit (Qiagen Inc., GmbH, Germany, USA). The quantity and quality of total RNA extracted were evaluated using Nanodrop 2000 spectrophotometer and Agilent Bioanalyzer. The sample was further subjected to cDNA library preparation to carry out Illumina sequencing [1].

RNA Seq library preparation and Illumina sequencing

The rRNA was removed from the total RNA using the Ribo-Zero rRNA removal kit (Illumina Inc., Singapore).The purified RNA sample was cut into short pieces using fragmentation buffer, and these fragments were used for first-strand cDNA synthesis using superscript II reverse transcriptase (Invitrogen, Carlsbad, Califonia, USA). Then, the second strand was synthesized and purified. Poly-A tailing and adapter ligation were done followed by cDNA synthesis for paired-end library preparation. Sequencing was performed with the help of NextSeq 500 using TruSeq v3-HS kit to generate 100bp paired-end reads (Illumina Inc., USA).

De novo transcriptome assembly and clustering

The quality of raw paired-end reads was assessed using FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and the adapter sequences and low-quality bases were removed using Cutadapt v3.5 and Sickle v1.33 tools, respectively [48,57]. The bases with Phred score ≥ 30 reads were retained, and the *de novo* assembler Trinity v2.13.2 was used for the transcriptome assembly, which combines three independent software modules: *Inchworm* assembles the data into full-length transcripts for a dominant isoform;*Chrysalis* clusters the contigs and constructs de Bruijn graphs for each cluster, representing the transcriptional complexity for a gene; *Butterfly* processes the graphs and reports full-length transcripts for alternatively spliced isoforms. CD-HIT v4.8.1 was used for the clustering to remove the redundancy of assembled contigs.

Assessment of gene completeness

An online tool TRAPID was used to analyse functional and comparative *de novo* transcriptome data set (http://bioinformatics.psb.ugent.be/trapid_02/), which compared the unigene transcripts against PLAZA 4.5 dicots plant database with an E-value <1E-5 for significant similarity search and annotation. Considering one or more hits in the TRAPID database for full length, quasi full length, or partial length based on ORF, unigene completeness was assessed.

Functional annotation and GO classification

A homology search using NCBI BlastX function against the non-redundant sequence database(nr), Gene Ontology (GO),and Kyoto encyclopedia of genes and genomes (KEGG) was carried out using Blast2GO software [59],where GO terms were classified based on three different functions that include biological process, molecular function, and cellular component. The KEGG Automated Annotation Server (KAAS) database was used to determine the biosynthetic pathways of various metabolites with an E-value of 1E-05.

SSR detection

Simple Sequence Repeats (SSRs) of *J.adhatoda* was identified using MISA tool (https://webblast.ipk-gatersleben.de/misa/) with default parameters to identify mono-, di-, tri-,tetra-, penta-, and hexa-nucleotide motifs as 10,6,5,5,5, and 5 repeats, respectively.

Gene expression analysis

The transcript quantification of *J.adhatoda* leaf was performed using the SALMON tool [58]. It quantifies the transcripts based on the GC content, which estimates the sensitivity and abundance of differential gene expression, and Quasi mapping of reads was done, where the transcripts were mapped itself fast and accurately.

Quantitative real-time PCR (qRT-PCR)analysis

To validate the expression of *de novo* assembled unigenes from *J.adhatoda*,qRT-PCR analysis was performed using QuantStudio 5 Real-Time PCR system (Thermo Scientific, Wilmington, Delaware, USA) and QuantiNova SYBR Green PCR Kit (Qiagen Inc., GmbH, Germany).Control reaction without a template was included for each primer set where glyceraldehyde 3-phosphate dehydrogenase (GAPDH) from *J.adhatoda* was used as an internal reference gene for estimation of gene expression. A comparative Ct ($2^{-\Delta\Delta Ct}$) method was used to analyse the qRT-PCR data, and fold change in gene expression was calculated using ΔCt values. The experiment was repeated with three technical and two biological replicates.

Results

RNA sequencing and *de novo* transcriptome assembly

RNA sequencing of leaf transcriptome from *J.adhatoda* generated 51,639,915 raw reads and the pre-processing of raw data was done to remove the adapter sequences, low-quality reads of 39,134,610 were generated, and a GC content of 41.69% was obtained.A total of 171,064 transcripts were assembled and the maximum and minimum unigene lengths were found to be 15630 bp and 201 bp, respectively with a N50 value is 2065 bp.

Assessment of gene completeness

Gene completeness of *J.adhatoda* revealed that the annotation of 30857 (28%) of full-length unigenes, 14071 (13.2%) quasi full-length unigenes, 22737 (21.3%) of partial unigenes, and 39221 (36%) unigenes did not match to any proteins in PLAZA 4.5 dicots plant database.

Functional annotation of unigenes

The assembled unigenes of *J. adhatoda* were annotated for sequence similarity search as well as for comparison using BLASTX against a non-redundant protein database at NCBI with a 1E-5. The results showed 105572 annotated unigenes and 40288 non-annotated unigenes, of which 4772 unigenes were predicted due to inadequate genomic

information in public databases. The unigene similarity search reveals that the plant has the best similarity with *Erythrantheguttata* (55%), *Coffeacanephora* (6.3%), *Utricula*riagibba (3.4%), among others. Thus, the results indicated that *J.adhatoda* is more closely associated with *Erythrantheguttata.*

Functional classification of unigenes

GO classification was done using the BLAST2GO tool in order to classify the genesbased on gene annotation into three different categories: molecular function, cellular component, and biological process and 47 sub-categories. A total of 143277 genes were assigned to 20 classes in the biological function, 26732 genes in organic substance metabolic process, 26713 genes in primary metabolic process, 21100 genes in nitrogen compound metabolic process, 19714 genes in the cellular metabolic process, and 11476 genes in the biosynthetic process. In molecular function, 93438 genes were classified into 14 sub-categories, which include the heterocyclic compound of 21276 genes, organic cyclic compound binding with 21276 genes, 16105 genes with transferase activity, 13563 genes with small-molecule binding, and 12904 genes with hydrolase activity. Additionally, 95725 genes were assigned to cellular component GO term with 13 sub-categories where intracellular GO term hadthe highest number of genes (23055), the intracellular part 21946 genes, 18296 genes of intracellular organelle, and 16066 genes of the membrane-bound organelle.

Biological pathway analysis

The biochemical pathways of *J.adhatoda* were identified by mapping the unigenes to the KEGG pathway database with the help of the KAAS and BLAST2GO software with 5981 unigenes annotated to 144 biochemical pathways. A total of 6409 unigenes were assigned to the metabolic pathway, including nucleotide metabolism, carbohydrate metabolism, and amino acid metabolism having 2539, 347, and 477 unigenes, respectively (Fig.3). The KEGG pathway analysis of secondary metabolite biosynthesis was divided into 12 sub-categories, where the highest number of unigenes were found in sesquiterpenoid and triterpenoid biosynthesis (57unigenes), followed by Ubiquinone and another terpenoid-quinone biosynthesis (26 unigenes), terpenoid backbone biosynthesis (22 unigenes), and phenylpropanoid biosynthesis (11 unigenes) (Fig.4).

Tryptophan biosynthesis:

The major genes involved in the tryptophan biosynthesis pathway from the KEGG database are presented in Table 2, and the genes identified in this pathway from transcriptome data havebeen depicted in Table 4.

Identification of simple sequence repeats (SSRs)

A total of 25978 SSRs were identified from 106886 sequences, with 138976200 bp of sequences containing 4374 sequences of more than 1 SSR, and 1951 number of SSRs were present in compound formation. The number of SSR loci with 8530 di-nucleotide repeats, 6802 tri-nucleotide repeats, 665 tetra-nucleotide repeats, 96 pentanucleotide repeats, and 61 hexanucleotide repeats are represented in Table 2.SSRs with five tandem repeats (3984) were the most common in *J. adhatoda,*followed by six tandem repeats (3902), seven tandem repeats (2289), nine tandem repeats (2087), eight tandem repeats (1622), and ten tandem repeats (589). Among di-nucleotide repeats, AT/AT was found to be the highest with 2968 repeats, followed by AG/CT with 2294 repeats, and tri-nucleotide repeats AAG/CTT has the highest frequency of 1737 and AAT/ATT with 1669 repeats and other motifs distributed uniformly (Fig. 5)

Transcript quantification

The expression levels of *de novo* assembled unigenes of the *J.adhatoda* leaf transcriptome were calculated based on TPM values with the help ofthe SALMON tool. In the tryptophan biosynthesis pathway, the TPM values of key enzymes involved are6.0903,33.6854, 11.527, 1.6959, and 8.1662for anthranilate synthase alpha, anthranilate synthase beta, arogenate/prephenate dehydratase, chorismate synthase, and chorismate mutase, respectively (Table 3). The top 10 most abundant unigenes in the *J.adhatoda* leaf transcriptome are represented in Table 4.

Gene expression analysis of *J.adhatoda* leaf tissue

The qRTPCR analysis was done to analyse the expression pattern of the selected tryptophan biosynthesis gene and validate the transcriptome data. The genes selected are anthranilate synthase alpha (EC: 4.1.3.27), anthranilate synthase beta (EC: 4.1.3.27), arogenate/prephenate dehydratase (EC: 4.1.1.48), chorismate synthase (EC: 4.2.3.5) and chorismatemutase (5.4.99.5). Anthranilatesynthase alpha, anthranilatesynthase, chorismate synthase and chorismate mutase showed significant up-regulation in mature leaf when compared to young leaf and root. Arogenate/prephenatedehydratase was downregulated in mature leaf tissue. Actin, Elongation factor 1 and Glyceraldehyde-3-phosphate dehydrogenase were used as a housekeeping gene for the gene expression analysis in this study. Anthranilate synthase beta subunit and arogenate/prephenate dehydratase with low TPM values were chosen for qRT PCR analysis,as they were also involved in the tryptophan biosynthesis pathway from the transcriptome data.

Discussion

The N50 value is the most important indices that were used to assess the quality of reads, which showed that high quality is suitable for intensive research [50,51,52,53]. Flavonoids are important secondary metabolites synthesised through the phenylpropanoid pathway. They are responsible for cytotoxicity and antioxidant potential through radical scavenging activity [18]. Chalconesynthase is the rate-limiting enzyme that catalyses the formation of pinocembrinchalcone fromCinnamoyl-CoA (EC: 2.3.1.74) in the flavonoid biosynthesis pathway [19]. The enzyme can also be found to produce polyketides thatare essential for plant metabolism and defense. These polyketides act as immunosuppressants, anti-cancer drugs, and insecticides [20,21]. Other key enzymes involved in the formation of flavanones, flavonols, and anthocyanidins are chalcone isomerise, anthocyanidinsynthase, dihydroflavonol-4-Reductase, and flavonolsynthase. These secondary metabolites are found to be potent hepatoprotective, anti-cancer,and anti-inflammatory agents [22]. Terpenoids are secondary metabolites that are also known as isoprenoids since they contain isoprenoid units. They are synthesised through two major metabolic pathways, namely, mevalonate pathway and the non-mevalonate pathway, which lead to the formation of sesquiterpenoids and monoterpenoids, respectively. 3-hydroxy-3-methylglutaryl-CoAreductase is the key regulating enzyme for sesquiterpenoid and triterpenoid biosynthesis pathways [23]. 1-deoxy-D-xylulose-5-phosphate synthase and 2C-methyl-D-erythritol 2,4-cyclodi-phosphate synthase are key regulatory enzymes required for the rate-limiting step of the non-mevalonate pathway leading to the biosynthesis of monoterpenoid compounds, which act as an anti-cancer agent [24]. Acridone is a major alkaloidal compound present in *J.adhatoda*,which acts as an anti-cancer agent [25]. It is aheterocyclic alkaloid that contains tricyclic rings with a carbonyl group at the ninth position and nitrogen at the tenth position [26]. There are various acridone derivatives such as glyforine, thioacridones, and acronycine, all of which exhibit pharmacological properties such as antimicrobial, antipsoriatic, anti-malarial, and anti-cancer activity [27,28,29,30]. Sesquiterpenoids are a large class of naturally occurring terpenoids containinga nitrogen group [39,40]. Phenylalanine is the major amino acid involved in the phenylpropanoid biosynthesis pathway that acts as a precursor. The amino acid is converted to cinnamic acid by the enzyme phenylalanine ammonia-lyase. This is the committed step of the biosynthetic pathway [54,55]. Tryptophan is an essential amino acid and substrate that is

required for the synthesis of serotonin, a neurotransmitter produced in the brain. It is also required for nicotinamide adenine dinucleotide (NAD+), which is an essential cofactor required for the regulation of serotonin and melatonin [33,34,35]. Tryptophan is converted to indole-3-pyruvate by aminotransferases from indole-3-acetic acid by the monooxygenases family of enzymes. This conversion plays a vital role in the auxin biosynthesis pathway [36,37,38]. The major genes identified in tryptophan biosynthesis were anthranilate synthase alpha (EC: 4.1.3.27)and anthranilate synthase beta (EC: 4.1.3.27). They are involved in the conversion of Chorismate into anthranilate;arogenate/prephenate dehydratase (EC: 4.1.1.48) converts 1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5-phosphate into (3-Indoyl)-glycerolphosphate, Chorismatesynthase (EC:4.2.3.5) converts chorismate into 5-O-(1-Carboxyvinyl)-3-phosphoshikimate,and Chorismatemutase (EC: 5.4.99.5) converts prephenate into chorismate. Anthranilate synthase is a rate-limiting enzyme involved in the terpenoidindole alkaloid pathway (TIA) [5]. This is a holoenzyme with two heterotetramers consisting of two alpha and beta subunits. The alpha subunits catalyse the formation of chorismate to anthranilate, which plays a vital role in the TIA pathway along with the binding site of tryptophan involved in feedback inhibition. The aminotransferases activity of the beta subunit plays a vital role in the transfer of amino groups from glutamine to the alpha subunit of anthrnilate synthase [10,11]. Chorismate is the final compound of the shikimate pathway that is formed by the enzyme chorismatesynthase. This enzyme catalyses the trans-1,4 elimination of phosphate from 5-enolpyruvylshikiukimate 3-phosphate, and it is of two types: fungal-type bifunctional chorismate synthase with NADPH-dependent flavinreductase and bacterial- and plant-type monofunctional chorismatesynthase [41,42]. Prephenateaminotransferase and arogenatedehydratase catalyse the final step of phenylalanine production [43,44,45]. Anthranilate synthase is a rate-limiting enzyme that catalyses chorismate to anthranilate in Indole pathway biosynthesis [5,10]. Anthranilate synthase consists of two subunits in its holoenzyme such as alpha and beta subunits. Alpha subunit plays a vital role in catalysing chorismate to anthranilate by feedback inhibition. The aminotransferase activity of the beta subunit of anthranilate synthasetransfers the amino group from glutamine to the alpha subunit [11]. Overexpression of anthranilate synthase increased the level of tryptophan that acts as a precursor for various biosynthetic pathways [12]. The production of vasicinone and vasicine was enhanced by stimulating anthranilate synthase activity by increasing the tryptophan and sorbitol in the culture [13].Chorismate mutase is a key enzyme that catalyses the formation of prephenate from chorismate for the biosynthesis of aromatic amino acids from the shikimate pathway [7]. In a study, Chorismate mutase was found to be a putative enzyme that is bifunctional and involved in the biosynthesis of phenylalanine, osmotic, and antibiotic tolerance [8,9].

Conclusion

This is the first study that has reportedon transcriptome assembly and annotation of the *J. adhatoda* and identified the uni-transcripts involved in the tryptophan biosynthesis pathway for its medicinal properties, as it acts as a precursor for acridone alkaloid biosynthesis and the validation of the key enzymes involved.Theseresults serve as an important resource for the molecular mechanism of various biosynthesis of medicinally important compounds from the plant.

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or non-profit sectors

Competing interests

The authors have no relavent financial or non-financial interests to disclose.

Author Contributions

Senthilkumar Palanisamy conceived the experimental study, guided the data analysis and reviewed the manuscript; Purushothaman Natarajan performed the omics box analysis; Deepthi Padmanabhan analysed the data,prepared figures and designed the manuscript;Adil Lateef collected the samples and performed the experiments. All authors contributed the manuscript at various stages.

Ethical approval

This is an observational study, hence no ethical approval is required.

References

1. AdilLateef, Sudheesh K Prabhudas & Purushothaman Natarajan (2018) RNA sequencing and *de novo* assembly of *Solanum trilobatum* leaf transcriptome to identify putative transcripts for major metabolic pathways. Scientific Reports. 8
2. Wangkhem Sandhya Chanu & Kananbala Sarangthem (2014) Phytochemical constituents of *Justicia Adhatoda* Linn. Found in Manipur. Indian Journal of Plant Sciences. 3(2), 19-21
3. Muhammad Gulfraz, Asma Ahmad, Muhammad Javaid Asad, Aliyasadiq, Umara Afzal, Muhammad Imran, Pervez Anwar, Asiya Zeenat, Kashif Sarfraz Abbasi, Sadaf Maqsood & Rahmat Ullah Qureshi (2011) Antidiabetic activities of leaves and root extracts of *Justicia adhatoda* Linn against alloxan induced diabetes in rats. African Journal of Biotechnology. 10(32), 6101-6106
4. Alka Sharma & Ashwani Kumar. Pharmacognostic studies on medicinal plants: *Justicia adhatoda* (2016) World Journal of Pharmaceutical Research. 5(7), 1674-1704
5. Jiayi Sun, Harish Manmathan, Cheng Sun & Chritie A. M (2016) Examining the transcriptional response of overexpressing anthranilate synthase in the hairy roots of an important medicinal plant *Cathranthusroseus* by RNA-seq. 16:108
6. Ercan Selcuk Unlu, OzgeKaya, Ismail Eker & Ekrem Gurel (2021) Sequencing, *de novo* assembly and annotation of *Digitalis ferruginea subsp. Schischkinii* transcriptome. Molecular Biology Reports. 48: 127-137
7. Kast P, Grisostomic C, Chen I.A, Li, S.L, Krengel U, Xue, Y.F (2000) A strategically positioned cation is crucial for efficient catalysis by chorismate mutase. Journal of Biology Chemistry. 275
8. Khanapur M, Alvala M, Prabhakar M, Kumar, K.S. Edwin, R.K. Saranya (2017) *Mycobacterium tuberculosis* chorismate mutase: A potential target for TB. Bioorganic and Medicinal Chemistry. 25: 1725-1736
9. Monti S. M, De Simone G & D'ambrosio. K (2016) L-histidinol dehydrogenase as a new target for old diseases. Current Topics in Medicinal Chemistry.16: 2369-2378
10. Hughes E H, Hong S B, Gibson S I, Shanks J V, San K Y (2004) Metabolite Engineering of the indole pathway in *Catharanthusroseus* hairy roots and increased accumulation of tryptamine and serpentine. Metabolite Engineering. 6(4):268-76
11. Peebles CA, Hong SB, Gibson SI, Shanks J V,San KY (2005) Transient effects of overexpressing anthranilate synthase alpha and beta subunits in *Catharanthusroseus* hairy roots. Biotechnology Programme. 21(5):1572-6
12. Ishihara A, Matsuda F, Miyagawa H, Wakasa K (2007) Metabolomics for metabolically manipulated plants: effects oftryptophan overproduction. Metabolomics. 3(3):319-34

13. Bharat Singh, Pooran M. Sahu, Ram A. Sharma (2017) Effect of elicitors on the production of pyrroloquinazoline alkaloids by stimulating anthranilate synthase activity in *Adhatoda vasica* Nees cell cultures. Planta. 246:1125-1137
14. G.Jayapriya & F.Gricilda Shoba (2015) Phytochemical analysis, antimicrobial efficacy and determination of bioactive components from leaves of *Justicia adhatoda* (Linn). Asian Journal of Plant Science and Research. 5(2), 43-51
15. G.Jayapriya&F. Gricilda Shoba (2015) GC-MS analysis of bio-active compounds in methanolic leaf extracts of *Justicia adhatoa* (Linn). Journal of Pharmacognosy and Phytochemistry. 4 (1), 113-117
16. ShikhaKalra, Bhanwar Lal Puniya, Deepika Kulshreshtha, Sunil Kumar, Jagdeep Kaur, Srinivasan Ramachandran & Kashmir Singh (2013) *De novo* Transcriptome sequencing reveals important molecular networks and metabolic pathways of the plant, *Chlorophytumborivilianum*. PLOS ONE. 8(12)
17. Dr. Subha Bose Banerjee & Kuntal Gupta (2021) *Justicia adhatoda* having multifunctional pharmacological properties can prevent COVID-19 disease. World Journal of Pharmaceutical and Medical Research. 7(9), 34-38
18. Ferreyra MLF, Rius SP, Casati P (2012) Flavonoids: biosynthesis, biological functionsand biotechnological applications. Frontier Plant Science. 3:222
19. BalachandranKarpaga Raja Sundari, RoliBudhwar, Bilikere S. Dwarakanth, S.P.Thyagarajan (2020) *De novo* transcriptome analysis unveals tissue-specific expression of candidate genes involved in secondary metabolite biosynthesis pathways of *Plumbagozeylanica*:implication for pharmacological potential. 3Biotech. 10:271
20. Dao TTH, Linthorst HJM, Verpoorte R (2011) Chalcone synthase and its function in plant resistance. Phytochemistry Review 10:397-412
21. Stewart CJR, Vickery CR, Bukart MD, Noel JP (2013) Confluence of structural and chemical biology: plant polyketide synthases as biocatalysts for a bio-based future. Current Opinion Plant Biotechnology. 16:365-372
22. Kumar S, Pandey AK (2013) Chemistry and biological activities of flavonoids: an overview. The Scientific World Journal
23. Tholl D (2015) Biosynthesis and biological functions of terpenoids in plants. Advance Biochemical Engineering/ Biotechnology. 148:63-106
24. Annadurai RS, Neethiraj R, Jayakumar V, Damodaran AC, Rao SN, Katta MA, Gopinathan S, Sharma SP, Senthilkumar V, Niranjan V (2013) *De novo* transcriptome assembly (NGS) of Curcuma longa L. rhizome reveals novel transcripts related to anticancer and anti-malarial terpenoids. PLOS ONE. 8:56217
25. RajkumarVeligeti, Rajesh BagepalliMadhu, JayashreeAnireddy, VisweswaraRaoPasupuleti, Vijaya Kumar Reddy Aluva, et al., (2020) Synthesis of novel cytotoxictetracyclicacridone derivatives and study of their molecular docking, ADMET, QSAR, bioactivity and protein binding properties. 10:20720
26. Heba A-EE-G, Farghaly AO, Mohammed AH (2018) Acridone-based acetyl-cholinesterase inhibitors;synthesis, antioxidant activity and molecular modeling. International Journal of Drug Delivery. 10:25-35
27. Bernaridone AMR, Azevedo AR, Pinheiro LCS, Rodrigues CR, Castro HC, Frugulhetti ICPP, Albuquerque MG (2008) SAR series of anti-HSV-1-acridone derivatives and a rational acridone-based design of a new anti-HSV-1 3H-benzo[b]pyrazolo[3,4-h]-1,6-napthyridine series. Bioorganic & Medicinal Chemistry. 16:313-21
28. Goodell JR, Madhok AA, Hiasa H, Ferguson DM (2006) Synthesis and evaluation of acridine-and acridone-based anti-hepes agents with topisomerase activity. Bioorganic & Medicinal Chemistry. 14:5467-80
29. Michael JP (2008) Quinoline, quinazoline and acridone alkaloids. Natural Products Reports. 25:166-87

30. Putic A, Stecher L, Prinz H, Muller K (2010) Structural-activity relationship studies of acridones as potential antisporatic agents.2.Synthesis and antiproliferative activity of 10 substituted hydroxyl-10H-acridin-9-ones against human keratinocyte growth. European Journal of Medicinal Chemistry. 45:5345-52
31. Hossain MT, Hoq MO (2016) Therapeutic use of *Adhatoda vasica*. Asian Journal of Medical and Bioological Research. 2(2):156-163
32. IfrahimIqbalChowdhury, Md. AtiarRahman, Mohammad A. Hashem, M. MosharefHossainBhuiyan, Dina Hajjar, Walla Alelwani, Arwa A. Makki et ., (2020) Supplements of aqueous combination of *Justicia adhatoda* and *Ocimumtenuiflorum* boost antioxidative effects and impede hyperlipidemia. Animal Models and Experimental Medicine. 3:140-151
33. Moffett JR, Namboodriri MA (2003) Tryptophan and the immune response. Immunology & Cell Biology. 81(4): 247-65
34. Ishihara A, Matsuda F, Miyagawa H, Wakasa K (2007) Metabolomics for metabolically manipulated plants:effects of tryptophan overproduction. Metabolomics. 3:319-334
35. Franz MarielleNogoy, Yu Jin Jung, Kwon-Kyoo Kang, Yong-Gu Cho (2019) Physio-chemical characterization and transcriptome analysis of 5-methyltryptophan resistant lines in rice. PLos ONE. 14(9)
36. Zhao Y (2012) Auxin biosynthesis: A simple two-step pathway converts tryptophan to indole-3 acetic acid in plants. Molecular plant. 5:334-338
37. Tao Y, Ferrer J.L, Ljung K, Pojer F, Hong F, Long J A, Li L, Moreno J E, Bowman M E, Ivans L J, et al., (2008) Rapid synthesis of auxin via a new tryptophan-dependent pathway is required for shade avoidance in plants. Cell. 133:164-176
38. Wei Wang, Yi Dai, Mingchum Wang, Wenpeng, Yang &Degang Zhao (2019) Transcriptome dynamics of double recessive mutant, o2o2o16o16, reveals the transcriptional mechanisms in the increase of its lysine and tryptophan content in Maize. Genes. 10:316
39. Vasas A, Hohmann J (2011) Xanthane sesquiterpenoids: structure, synthesis and biological activity. Natural product Reseach. 28:824-842
40. De-Li-Chen, Bo-Wen Wang, Zhao-cui Sun, Jun-Shan Yang, Xu-Dong Xu&Guo-XU Ma (2020) Natural nitrogenous Sesquiterpenoids and their Bioactivity: A Review. Molecules. 25:2485
41. Schaller A, Vanafferden M, Windhofer V, Bulow S, Abel G, Schmid J et al., (1991) Purification and characterization of chorismate synthase from *Eugelenagracius*-comparision with chorismate synthases of plant and microbial origin. Plant Physiology. 97:1271-1279
42. Maeda H & Dudareva N (2012) The shikimate pathway and aromatic amino acid biosynthesis in plants. Annual review of plant biology. 63:73-105
43. Cho MH, Corea ORA, Yang H, Bedgar DL, Laskar DD, Anterola AM et al., (2007)Phenylalanine biosynthesis in *Arabidropsis thaliana*- identification and characterization of arogenate dehydratases. Journal of Biological Chemistry. 282:30827-30835
44. Huang T, Tohge T, Lytovchenko A, Fernie A R &Jander G (2010) Pleiotropic physiological consequences of feedback insensitive phenylalanine biosynthesis in *Arabidopsis thaliana*. The Plant Journal. 63:823-835
45. Takayuki Tohge, Mutsumi Watanbe, Rainer Hoefgen & Alisdair R. Fernie (2013) Shikimate and phenylalanine biosynthesis in the green lineage. Frontiers in Plant Science. 12
46. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al., (2011) Full length trasnscriptome assembly from RNA-Seq data without a reference genome. Natural Biotechnology. 29:644-652

47. Zhao QY,Wang Y, Kong YM, Luo D, Li X, Hao P. (2011) Optimizing *de novo* transcriptome assembly from short-read RNA Seq data: A comparative study. BMC Bioinformatics. 12
48. Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Natural Reviews Genetics. 12:671-682
49. HuLS,Hao CY, Fan R, Wu BD, Tan LH, Wu HS et al., (2015) *De novo* assembly and Characterization of fruit Transcriptome in Black pepper (*Piper nigrum*). PLos ONE. 10
50. Shi Y, Yan X, Zhao PS, Yin HX, Zhao X, Xiao HL et al., (2013) Transcriptome analysis of tertiary relict plant, Extreme Xerophyte *Reeaumuriasoongorica* to identify genes related to drought adaptation. PLos ONE. 8
51. Long Y, Zhang JW, Tian XJ, Wu SS, Zhang , Zhang J, et al., (2014) *De novo* assembly of the desert tree *Haloxylonammodendron* (CA Mey). Based on RNA Seq data provides insight into drought response, gene discovery and marker identification. BMC Genomics. 15:1111
52. Feng C, Chen M, Xu CJ, Bai L, Yin XR, Li X et al., (2012) Transcriptomic analysis of Chinese bayberry (*Myricarubra)* fruit development and ripening using RNA Seq. BMC Genomics. 13:19
53. Dian-YunHou, Lin Chun Shi, Meng-Meng Yang, Jiong LI, ShuangZho, Hong-Xiao Zhang, Hua-Wei Xu (2018) Transcriptomic analysis of leaf and fruit of *Cornusofficinalis* using Illumina platform. PLos ONE
54. Shang QM, Li L, Dong CJ (2012) Multiple tandem duplication of phenylalanine ammonia lyase genes in *Cyrcumissativus* L. Planta 236(4):1093-105
55. Xiaowei Wang, Hong yan Hu, Zhijun Wu, Haili Fan, Guowei Wang, Tuanyao Chai & Hong Wang (2021) Tissue-specific transcriptome analyses reveal candidate genes for stilbene, flavonoid and anthraquinone biosynthesis in medicinal plant *polygonumcuspidatum*. BMC Genomics. 22:353
56. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMB net.journal. 17(1):10-12
57. Joshi N.A &Fass J.N (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files.
58. Rob Patro, Geet Duggal, Michael I Love, Rafeal A Irizarry & Carl Kingsford (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nature methods. 14: 417-419
59. Ana Conesa & Stefan Gotz (2008) Blast2GO: A Comprehensive suite for functional Analysis in Plant Genomics. International Journal of Plant Genomics. 12

Tables

Table 1 Summary of Illumina paired-end sequencing and *de novo* assembly of *J. adhatoda* transcriptome

Particulars	Numbers
Number of raw reads	51,639,915
Number of clean reads	39,134,610
No. of bases (after processing)	33,247,241
Mean Phred score	37.45
Total transcripts	171,064
Percentage of successful assembly from raw reads	64.38
Average length (transcripts)	1383.89
Median contig length	1096
GC %	41.69
Contig N50 (transcripts)	2065

Table 2 Major enzymes involved in Tryptophan biosynthesis pathway (https://www.genome.jp/pathway/map00400+M00023).

Enzyme name	E.C.Number
Anthranilate synthase	4.1.3.27
Anthranilate phosphoribosyltransferase	2.4.2.18
Phosphoribosylanthranilate isomerase	5.3.1.24
Indole-3-glycerol phosphate synthase	4.1.1.48
Tryptophan synthase alpha chain	4.2.1.20

Table 3 Distribution and frequency of EST-SSRs identified in *Justicia adhatoda*

Motif length	Repeat numbers							
	5	6	7	8	9	10	>10	Total
di	0	2223	1443	1028	2009	468	1359	8530
tri	3397	1520	806	570	76	115	318	6802
tetra	461	145	27	24	2	4	2	665
penta	80	7	5	0	0	2	2	96
hexa	46	7	8	0	0	0	0	61
Total	3984	3902	2289	1622	2087	589	1681	

Table 4 Major genes involved in tryptophan biosynthesis from the *J.adhatoda* leaf transcriptome.

Gene name	E C Number	Unigene ID	Unigene length	TPM value	No. Of Unigene
Anthraniltesynthase alpha	4.1.3.27	TRINITY_DN13973_c0_g1_i2	524	6.0903	1
Anthranilate synthase beta	4.1.3.27	TRINITY_DN9855_c0_g1_i4	2058	33.6854	1
Arogenate/Prephenatedehydratase	4.1.1.48	TRINITY_DN8030_c0_g1_i1	1605	6.9556	7
		TRINITY_DN11169_c0_g1_i1	2648	0.8030	
		TRINITY_DN11169_c0_g1_i2	3676	2.3483	
		TRINITY_DN11169_c0_g1_i3	3139	8.1132	
		TRINITY_DN9715_c0_g1_i2	1765	11.527	
		TRINITY_DN9715_c0_g1_i3	666	6.4298	
		TRINITY_DN13484_c0_g1_i1	769	5.9688	
Chorismatesynthase	4.2.3.5	TRINITY_DN12680_c0_g1_i3	2687	1.6959	2
		TRINITY_DN12680_c0_g1_i5	1132	0.8975	
Chorismatemutase	5.4.99.5	TRINITY_DN11294_c0_g1_i1	2184	8.1662	5
		TRINITY_DN11294_c0_g1_i5	1293	2.2369	
		TRINITY_DN11294_c0_g1_i6	2287	2.1131	
		TRINITY_DN11294_c0_g1_i7	2016	1.0050	
		TRINITY_DN11295_c0_g1_i1	1170	4.4995	

Table 5 Ten most abundant genes from the leaf tissue of *J.adhatoda.*

Transcript ID	Gene Name	TPM
TRINITY_DN5212_c0_g1_i1	Nuclear poly(A) polymerase 3-like	5630.023
TRINITY_DN1643_c0_g1_i1	Formin-like protein 20	729.4972
TRINITY_DN11570_c0_g1_i1	uncharacterized protein LOC105178309	2639.996
TRINITY_DN1283_c0_g1_i1	uncharacterized protein LOC111020109	1865.95
TRINITY_DN4476_c0_g1_i1	Zinc finger MYM-type protein 1-like	1572.039
TRINITY_DN959_c0_g1_i1	Sugar transporter ERD6-like 6	1278.84
TRINITY_DN7445_c0_g1_i1	ABC transporter B family member 11-like	1211.744
TRINITY_DN5489_c0_g1_i1	Ethylene-responsive transcription factor TINY-like	1066.038
TRINITY_DN448_c0_g1_i1	PREDICTED: uncharacterized protein LOC109132446	1059.974
TRINITY_DN10357_c0_g1_i6	Methyl-CpG-binding domain-containing protein 13-like isoform X3	1017.646
TRINITY_DN2853_c0_g1_i1	Protein trichome birefringence-like 19	1007.04

Figures

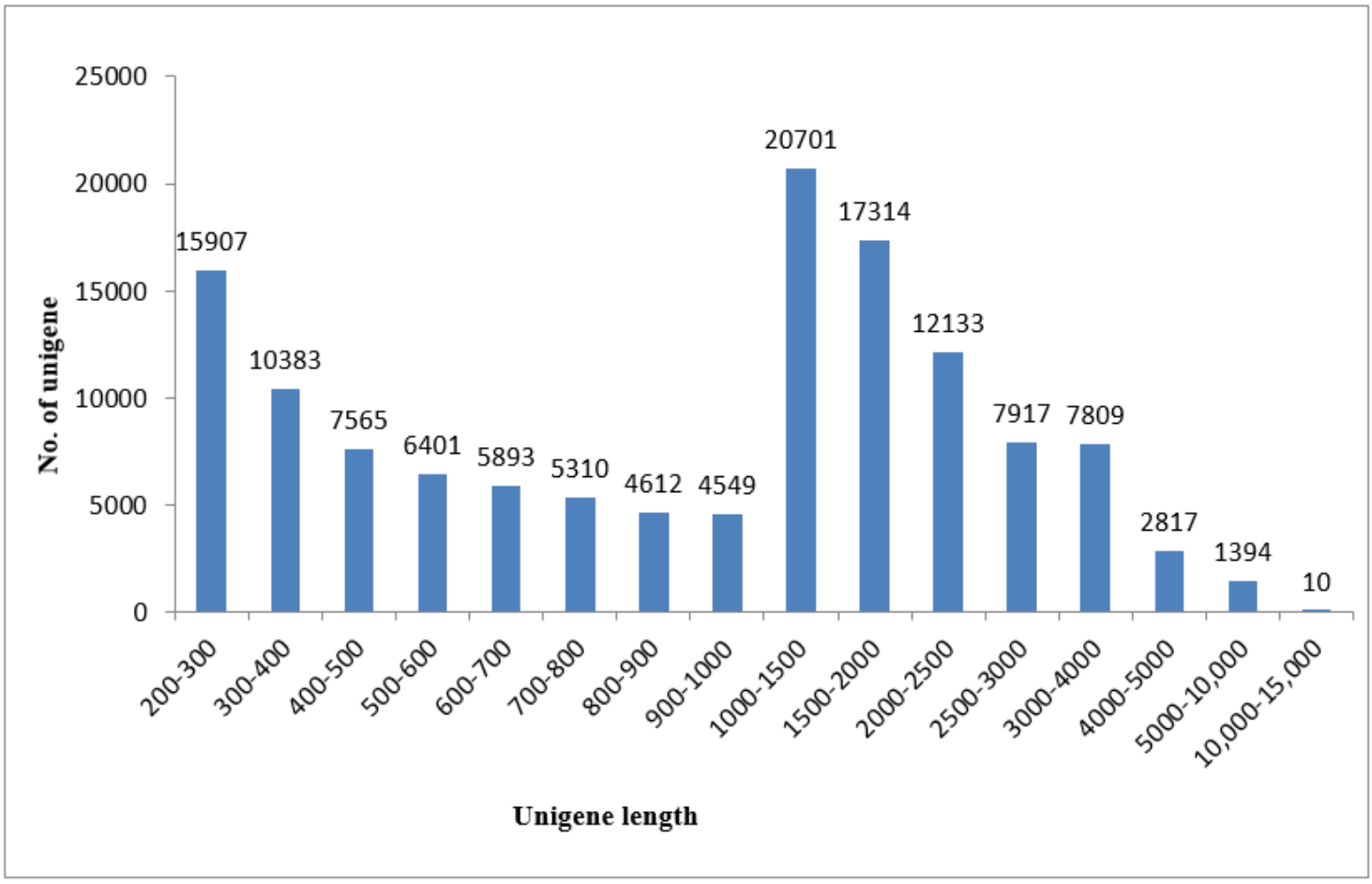

Figure 1

Unigene length distribution of *J. adhatoda* leaf transcriptome

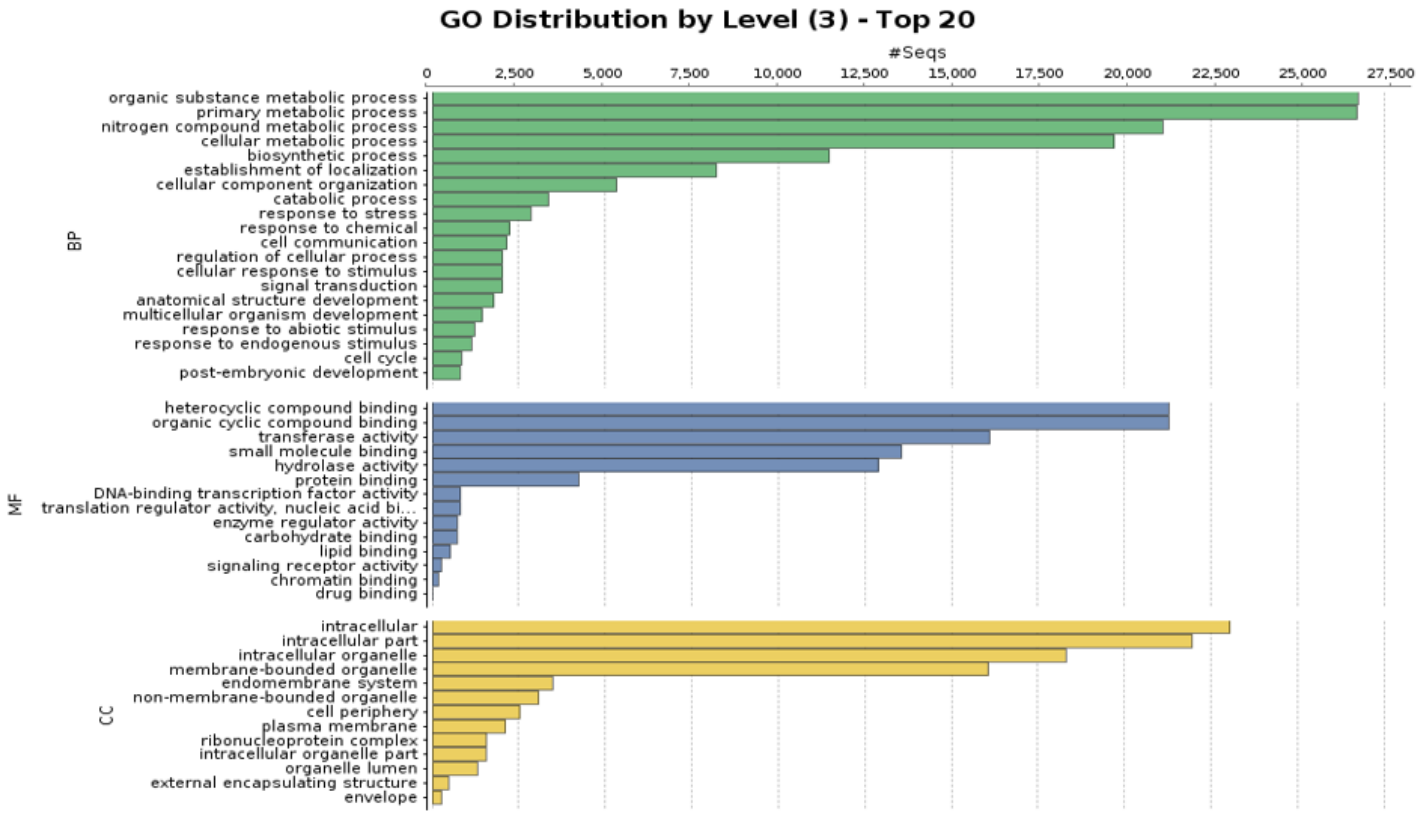

Figure 2

Gene Ontology classification of *J. adhatoda* into three major categories i.e., cellular component, molecular function, and biological process

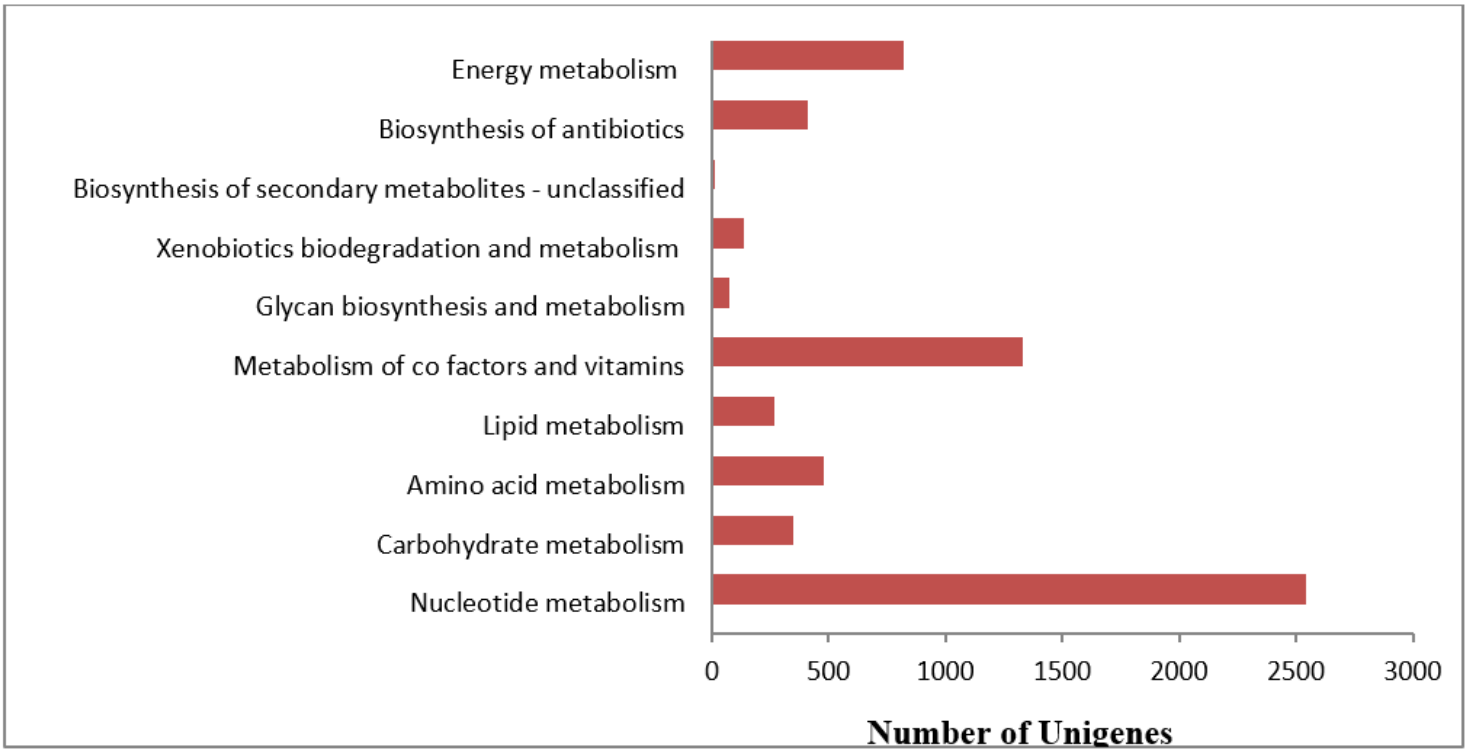

Figure 3

KEGG pathway of unigene based on the metabolic pathway

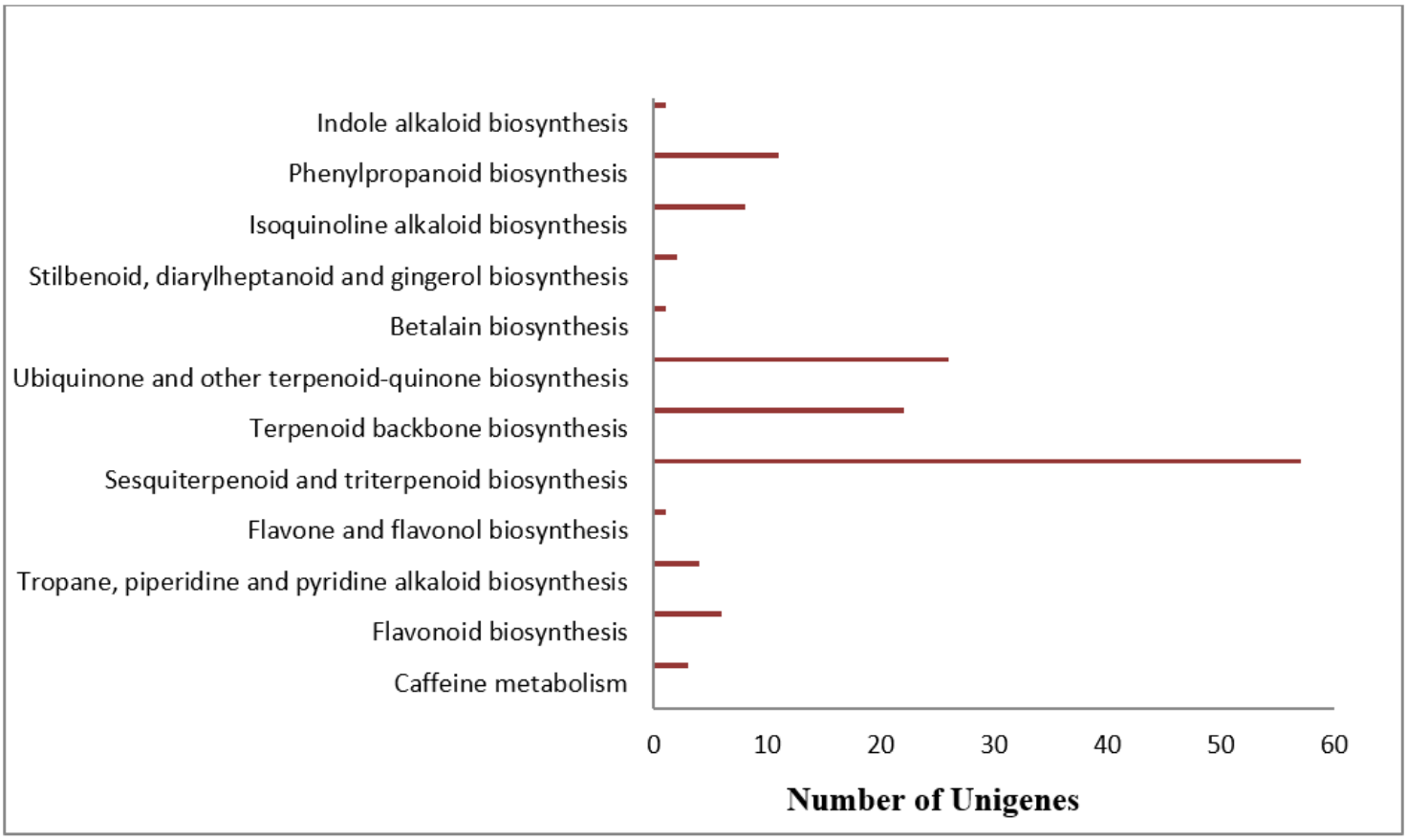

Figure 4

KEGG pathway of unigene based on secondary metabolite biosynthesis

Figure 5

Tryptophan biosynthesis pathway from KEGG database that represents the various enzymes identified from transcriptome data concerning *J. adhatoda* in each color along with its enzyme code (EC number). The KEGG pathway map is adapted from map00400 https://www.kegg.jp/keggbin/highlight_pathway?scale=1.0&map=map00400&keyword=tryptophan.

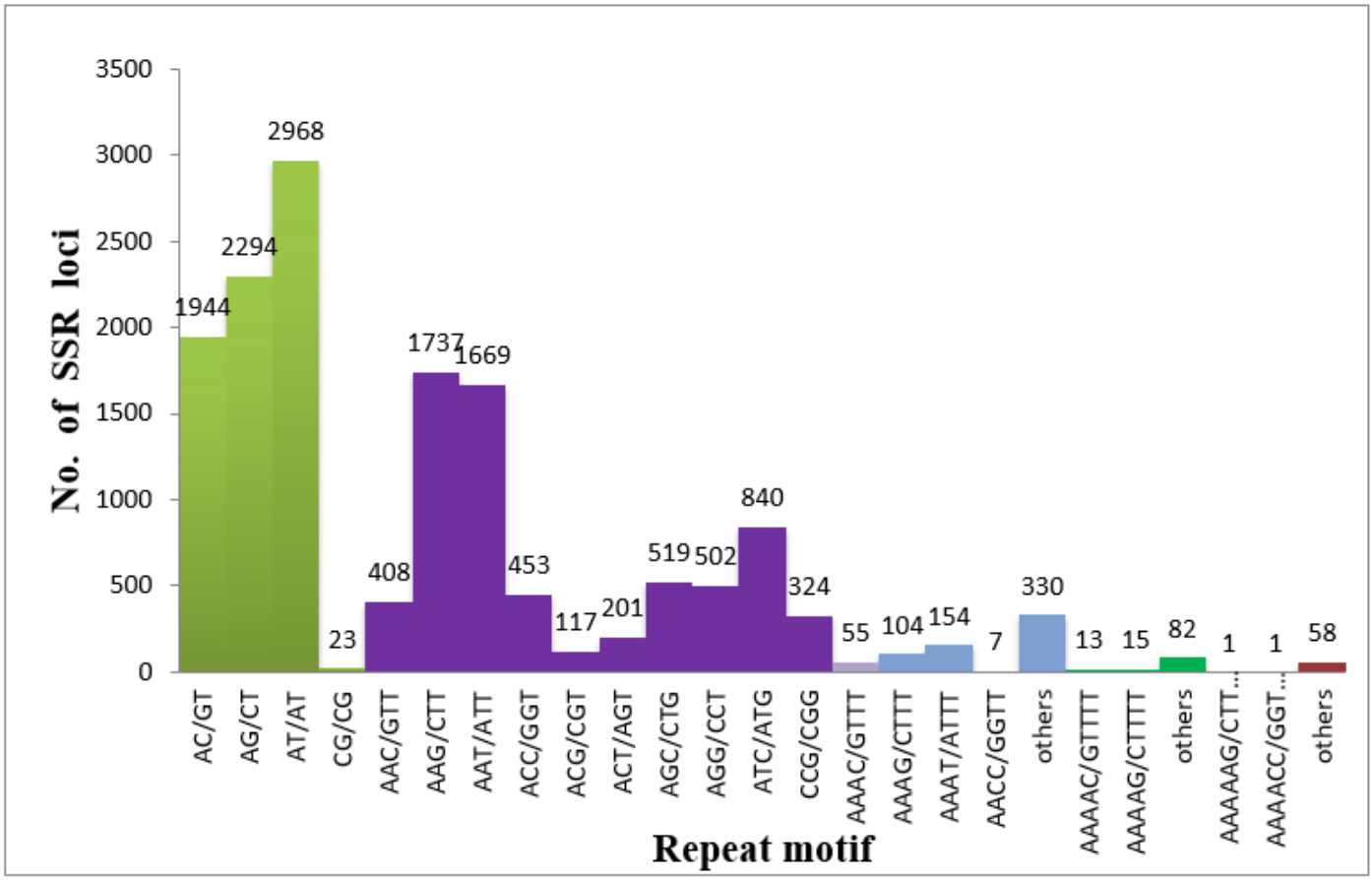

Figure 6

Summary of SSR types identified in *J.adhatoda* leaf transcriptome

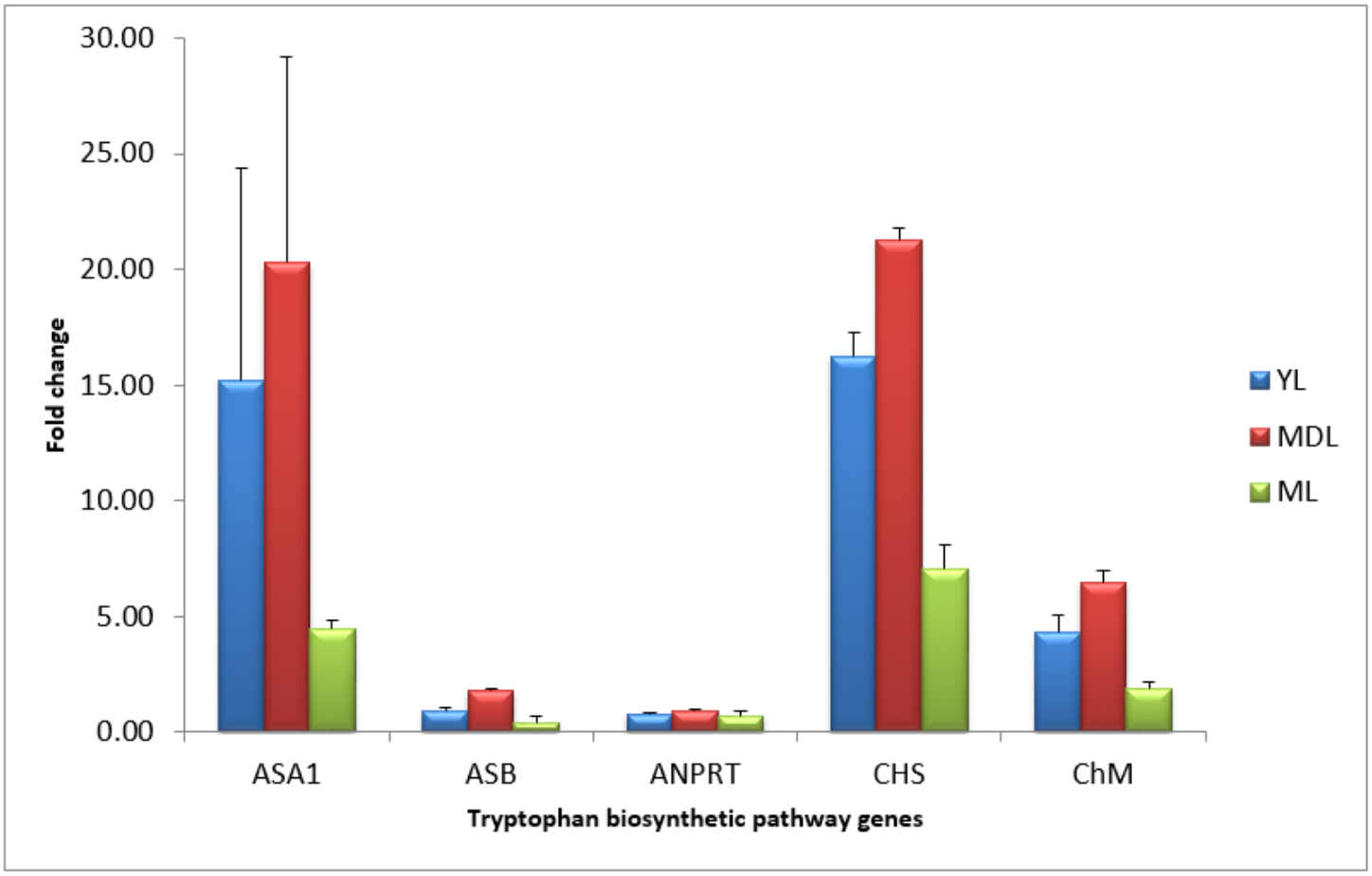

Figure 7

qRT-PCR validation of anthranilate synthase alpha 1 (ASA1), anthranilate synthase beta (ASB), arogenate/prephenate dehydratase (ANPRT), chorismate synthase and chorismate mutase in a young leaf (YL), mid-mature leaf (MDL), and root (ML) of *J. adhatoda*.