Canadian Science Publishing | Genome

Identification of full-length genes involved in the biosynthesis of β-caryophyllene and Lupeol from the leaf transcriptome of Ayapana triplinervis

Journal:	*Genome*
Manuscript ID	gen-2024-0005.R1
Manuscript Type:	Note
Date Submitted by the Author:	28-May-2024
Complete List of Authors:	Parani, Madasamy; SRM Institute of Science and Technology, Department of Genetic Engineering Tanuja, Tanuja; SRM Institute of Science and Technology, Department of Genetic Engineering
Is the manuscript for consideration in a Special Issue or Collection?:	Not applicable (regular submission)
Keyword:	β-caryophyllene, Eupatorium ayapana, Eupatorium triplinervis, lupeol, transcriptome

Identification of full-length genes involved in the biosynthesis of β-caryophyllene and Lupeol from the leaf transcriptome of *Ayapana triplinervis*

Tanuja and Madasamy Parani*

Department of Genetic Engineering, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur 603203, Kanchipuram, Chennai, TN, India

*Author for correspondence: paranim@srmist.edu.in

Abstract

β-Caryophyllene possesses potential anticancer properties against various cancers, including breast, colon, and lung cancer. Therefore, the essential oil of *Ayapana triplinervis*, which is rich in β-caryophyllene, can be a potential herbal remedy for treating cancer. However, molecular and genomic studies on *A. triplinervis* are still sparse. In this study, we obtained 14.7 Gb of RNA-Seq data from *A. triplinervis* leaf RNA and assembled 1,37,554 transcripts with an N50 value of 1,437 bp. We annotated 72,436 (52.7%) transcripts and mapped 10,640 transcripts to 156 biochemical pathways. Among them, 218 were related to terpenoid backbone biosynthesis, while 27 were linked to sesquiterpenoid and triterpenoid pathways. Ninety-four transcripts were annotated in the β-caryophyllene and lupeol pathways. From these transcripts, for the first time, we identified 25 full-length genes encoding all the 17 enzymes involved in β-caryophyllene biosynthesis and an additional five genes involved in lupeol biosynthesis. These genes will be useful for the metabolic engineering of β-caryophyllene and lupeol biosynthesis, not just in *A. triplinervis* but also in other species.

Keywords: β-caryophyllene, *Eupatorium ayapana*, *Eupatorium triplinervis*, lupeol, transcriptome

INTRODUCTION

A*yapana triplinervis* (synonyms: *Eupatorium triplinerve* and *Eupatorium ayapana*), known as Ayapani or Ayapana, belongs to the Asteraceae family. Studies using leaf extracts and cell lines provided compelling evidence that it contains potential anticancer compounds (Arung et al. 2012, Suja. et al. 2019, Gayathri and Madhan Shankar 2012). The essential oil from *A. triplinervis* leaves contains up to 19.7% β-caryophyllene (Gupta et al. 2004, Mala et al., 1999) which has potential anticancer properties against a variety of cancers, including breast, colon, lung, glioblastoma, myeloma, and melanoma (Yang et al. 2015; Chung et al. 2019; Irrera et al. 2020; Dahham et al. 2021; Mannino et al. 2021).

lupeol is another potential anticancer compound against various cancers, such as breast, cervical, retinoblastoma, osteosarcoma, and prostate (Prasad et al. 2018; Hsu et al. 2019; Zhang et al. 2022; Che et al. 2022; Maurya et al. 2022). Therefore, Identifying the genes involved in caryophyllene and lupeol biosynthesis in *A. triplinervis* will be useful. However, the dearth of genomic data makes gene discovery challenging. Therefore, we assembled the *A. triplinervis* leaf transcriptome with the aim of identifying the full-length genes involved in the synthesis of β-caryophyllene and lupeol.

METHODOLOGY

Plant collection and RNA isolation

A. triplinervis plants were collected from Kannur, located in Kerala, India. These plants were grown in a greenhouse, maintaining 25/18°C day/night temperatures, a light/dark duration of 16/8 hours, and a relative humidity between 60 and 70%. We isolated RNA from the leaves using RNAiso Plus reagent (Takara, Japan), treated with DNase I, and purified using the RNeasy MinElute Clean-up Kit (Qiagen, Germany). We used the total RNA with an OD260/OD280 ratio of 1.9 and an RNA integrity number of 8.5 for cDNA synthesis.

cDNA synthesis and library preparation

The cDNA library was prepared using 1 µg total RNA and the Illumina TrueSeq mRNA library kit (Illumina, USA) as per the manufacturer's protocol. The cDNA quality and quantity were determined using the Bioanalyzer (Agilent Technologies, USA) and the Qubit dsDNA HS kit and fluorimeter. The library was then sequenced using the Illumina NextSeq500 system (Illumina, USA).

De novo assembly

The data quality was assessed using FastQC (Andrews 2010). The reads with quality scores <30 were filtered using Trimmomatic (Bolger et al. 2014). Subsequently, it assembled using the Trinity software using default parameters (K-mer 25, overlap length 12 - 24 bp, and overlap identity 100%) (Grabherr et al. 2011). Shorter and redundant sequences were removed using the CD-HIT-EST software (Li and Godzik 2006).

Functional annotation and pathway identification

The assembled transcriptome was annotated using the BLASTx tool and the NCBI non-redundant (https://ftp.ncbi.nlm.nih.gov/blast/db/) with an E-value cut-off of $1E^{-5}$. Annotation was performed using NR, Uniport, Blast2GO, and KEGG. The transcripts related to metabolic pathways were identified using the KEGG pathway analysis.

Expression analysis by RT-qPCR

The expression of FPPS1, FPPS2, SQS, SQS2, SQE2, CS, LS, and SE1 genes in the leaves was studied by RT-qPCR using the elongation factor 1 alpha gene as the internal reference. Primers were designed from the 3' UTR region using the Prime3Plus software (https://www.primer3plus.com/), and the details are provided in Table S1. The RT-qPCR reactions in triplicate contained 50 ng cDNA, 1X TB Green Premix Ex Taq II (Takara, Japan),

0.25 mM primers, 1X ROX dye, and nuclease-free water. The thermal cycling used was 95°C for 5 minutes, followed by 40 cycles at 95°C for 10 seconds and 65°C for 20 seconds.

RESULTS

Transcriptome assembly and annotation

We obtained 98 million 150 bp reads (available at NCBI under accession number SRR24848950 and Bioproject ID PRJNA980972), and 80 million quality-filtered reads with Phred score ≥Q30. We assembled 1,37,554 transcripts with an average length of 879 bp and an N50 value of 1,437 bp. Among them, 72,436 transcripts (52.7%) were annotated, which included 41,232 transcripts (60%) aligned with established gene models and 31,204 transcripts (40%) showing similarity with hypothetical, probable, putative, or uncharacterized proteins.

Genes involved in β-Caryophyllene and lupeol biosynthesis

Farnesyl pyrophosphate (FPP) is the common precursor for β-caryophyllene and lupeol. It is synthesized from acetyl-CoA and D-glyceraldehyde-3-phosphate by the action of sixteen enzymes. From FPP, β-caryophyllene is formed by a single enzyme, while lupeol synthesis requires three enzymes. Our analysis revealed 94 transcripts coding for all the 20 enzymes needed for β-caryophyllene and lupeol biosynthesis. Of these, 30 were full-length genes. The length of the coding sequences for these genes ranged from 819 bp to 2,277 bp, translating to protein lengths of 273 to 759 amino acids. Three genes were identified for the 1-deoxy-D-xylulose 5-phosphate synthase (DXS) enzyme. The enzymes such as acetyl-CoA acetyltransferase (AACT), hydroxymethylglutaryl-CoA reductase (HMGR), phosphomevalonate kinase (PMK), and 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) had two full-length genes each. Hydroxymethylglutaryl-CoA synthase (HMGS), mevalonate kinase (MVK), β-caryophyllene synthase (CS), and lupeol synthase (LS) were each encoded by a single gene. All the gene sequences were deposited to the NCBI GenBank under

the accession numbers OR544240 to OR544270 (Table 1) and the nucleotide and protein sequences are given in Supplementary files S2 and S3 respectively.

Gene expression analysis by RT-qPCR

We analyze the expression of FFPS1, FPPS2, SQS1, SQS2, SE2, CS, LS, and SE1 genes by RT-qPCR, using the elongation factor 1-alpha gene as a reference. All samples were analyzed in triplicate. The expression levels in leaves were ascertained from the average cycle of quantification (Cq), which ranged between 18.2 and 26.6. For three enzymes, we analyzed two genes for each enzyme. In these cases, the two genes showed differential expression: FPPS1>FPPS2 (23 fold), SQS1>SQS2 (37 fold), and SES2>SES1 (19 fold) (Figure 1).

Discussion

A. triplinervis is a significant medicinal plant with potential anticancer properties. However, there is a dearth of genetic and molecular studies on this species. In this study, we assembled the leaf transcriptome and identified the full-length genes involved in the biosynthesis of two potential anticancer compounds, β-Caryophyllene and lupeol. These two compounds are derived from the common precursor, farnesyl pyrophosphate (FPP), which is synthesized through the mevalonic acid (MVA) pathway in the cytosol and the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway in the plastids (Vranová et al. 2013). The MVA pathway produces isopentyl pyrophosphate (IPP) from acetyl-CoA through a sequence of reactions involving acetyl-CoA C-acetyl transferase (AACT), 3-hydroxy-3-methylglutaryl coenzyme A synthase (HMGS), 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR), mevalonate kinase (MVK), 5-phosphomevalonate kinase (PMK), and mevalonate diphosphate decarboxylase (MDD). We identified ten genes that encode these seven enzymes. We identified two full-length genes for AACT, namely AtAACT1 and AtAACT2. Two homologous genes, HaAACT1 and HaAACT2, were reported from *Helianthus annuus*. The protein similarity

between AtAACT1 and HaAACT1 was 90%, whereas AtAACT2 showed a 97% similarity with HaAACT2 (Huang et al. 2023). We also identified two genes coding for PMK that showed 81% protein similarity, indicating that they are distinct genes. In addition, we have discovered single genes for HMGR, HMGS, MVK, and MDD.

In the MEP pathway, dimethylallyl pyrophosphate (DMAPP) is synthesized from glyceraldehyde 3-phosphate, involving 1-deoxy-D-xylulose 5-phosphate synthase (DXS), 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS), 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase (HDS), and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR). We identified 11 full-length genes that encode all of these enzymes. Three genes (AtDXS1, AtDXS2, and AtDXS3) were discovered for DXS, each having unique coding sequences. DXS1 is mostly linked to primary metabolism, whereas DXS2 has a crucial function in secondary metabolism (Walter et al. 2002; Phillips et al. 2007). DXS3 is involved in the synthesis of essential chemicals crucial for plant survival (Cordoba et al. 2011). Three DXS genes (SrDXS1, SrDXS2, and SrDXS3) were identified from *S. rebaudiana* (Kim et al. 2015), which shared 95 to 97% similarity with AtDXS1, AtDXS2, and AtDXS3. We identified single genes for MCT, CMK, MCS, and HDS. In the MEP pathway, IPP can be produced from hydroxy-2-methyl-2-butyl-4-phoshate by 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR) and DMAPP by isopentenyl diphosphate isomerase (IPPI). We identified two full-length genes for HDR (AtHDR1 and AHDR2) and IPPI (AtIPPI1 and AtIPPI2). Similar HDR and IPPI genes were found in the transcriptome of *S. rebaudiana*, and SrIPPI1 and SrIPPI2 genes were expressed in the leaf and trichome (Kim et al. 2015).

Condensation of DMAPP and IPP by Geranyl diphosphate synthase (GPPS) results ingeranyl pyrophosphate (GPP), which is further condensed with IPP by farnesyl pyrophosphate synthase

(FPPS) to produce farnesyl pyrophosphate (FPP). We found one GPPS gene with a 1,212 bp CDS encoding 404 amino acids. We identified two FPPS genes (AtFPPS1 and AtFPPS2) with the same CDS lengths (1,026 bp), but differing sequences as indicated by 89% and 95% similarity at nucleotide and amino acid levels, respectively. The distinctness of these two genes predominantly arises from the amino acid composition in the N-terminal region. The conversion of FPP into β-caryophyllene is catalyzed by caryophyllene synthase (CS). We identified a CS gene in *A. triplinervis* (AtCS1) that showed 79% similarity to the same gene found in *Artemisia annua* (AaCS). Transient expression of AaCS in *Nicotiana benthamiana* leaves produced 27.5 mg of β-caryophyllene per kg of fresh weight (Muthusamy et al. 2020). Transfer of AaCS1 gene to Synechocystis and *Rhodobacter capsulatus* produced 46.4 µg/L and 139 mg/L β-caryophyllene, respectively (Reinsvold et al. 2011 and Hilgers et al. 2021). Because β-caryophyllene is naturally abundant in *A. triplinervis*, its CS gene might enhance the β-caryophyllene content more efficiently in other organisms as well.

Two FPP molecules can combine to produce squalene, which is then converted to lupeol. Currently, the lupeol content in *A. triplinervis* has not been determined. However, we identified five genes that encode the three enzymes necessary for lupeol synthesis from FPP. In the first step, FPP gets converted into squalene by squalene synthase (SQS). From our transcriptomic analysis, we have successfully identified two genes for SQS, AtSQS1 and AtSQS2. Recently, AlSQS1 and AtSQS2 from *Atractylodes lancea* were functionally characterized, and found to be expressed more in stems than in leaves and rhizomes (Wu et al. 2022). AtSQS1 showed a 95% similarity to AlSQS1, while AtSQS2 exhibited an 89% similarity with AlSQS2. AtSQS1 had a higher expression level compared to AtSQS2 in *A. triplinervis*. In the next step, squalene is oxidized to squalene-2,3-epoxide by squalene epoxidase (SE). We identified two SE genes, AtSE1 and AtSE2, with 79% similarity, signifying their distinctiveness. PgSE1 and PgSE2 were found to be involved in ginsenoside

biosynthesis in *Panax ginseng* (Han et al. 2010). In the final step, squalene-2,3-epoxide is transformed into lupeol by lupeol synthase (LS). We identified an AtLS gene (2,277 bp CDS; 579 amino acids) with 94% similarity to the TkLS gene from *T. koksaghyz*. Significantly, heterologous expression of TkLS in yeast showed a 127-fold increase in lupeol content (Bröker et al. 2018).

This study significantly expanded our knowledge regarding the genetic resource of *A. triplinervis*, discovering full-length genes encoding all the enzymes required for β-caryophyllene and lupeol synthesis. For the first time, we identified 25 genes associated with β-caryophyllene biosynthesis and five genes related to lupeol biosynthesis. These genes are crucial for the metabolic engineering of β-caryophyllene and lupeol biosynthetic pathways, aimed at increasing metabolite yield.

Competing interest:

We have no conflicts of interest to disclose.

Author Contributions: The following statements should be used: Conceptualization, MP, and Tanuja; methodology, Tanuja; investigation, Tanuja; writing-original draft preparation, Tanuja; writing-review and editing, MP; supervision, MP; project administration, MP.; funding acquisition, MP. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the SRM-DBT Partnership Platform for Contemporary Research Services and Skill Development in Advanced Life Sciences Technologies (Order No. BT/PR12987/INF/22/205/2015).

Data Availability Statement: The SRA data was submitted to the NCBI with the Bio project ID PRJNA980972. Full-length gene was submitted under the GenBank accession numbers OR544240 to OR544270.

REFERENCES:

Arung TE, Kuspradini H, Kusuma WI, et al (2012) Arung_2012_Validation of Eupatorium triplinerve Vahl Leaves, a Skin Care Herb from East Kalimantan, Using a Melanin Biosynthesis Assay _ Elsevier Enhanced Reader. J Acupunct Meridian Stud 5:87–92

Andrews S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bröker JN, Müller B, van Deenen N, et al (2018) Upregulating the mevalonate pathway and repressing sterol synthesis in Saccharomyces cerevisiae enhances the production of triterpenes. Appl Microbiol Biotechnol 102:6923–6934. https://doi.org/10.1007/s00253-018-9154-7

Che S, Wu S, Yu P (2022) Lupeol induces autophagy and apoptosis with reduced cancer stem-like properties in retinoblastoma via phosphoinositide 3-kinase/ protein kinase B/mammalian target of rapamycin inhibition. Journal of Pharmacy and Pharmacology 74:208–215. https://doi.org/10.1093/jpp/rgab060

Chung KS, Hong JY, Lee JH, et al (2019) B-Caryophyllene in the essential oil from chrysanthemum boreale induces G1 phase cell cycle arrest in human lung cancer cells. Molecules 24:. https://doi.org/10.3390/molecules24203754

Cordoba E, Porta H, Arroyo A, et al (2011) Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize. J Exp Bot 62:2023–2038. https://doi.org/10.1093/jxb/erq393

Dahham SS, Tabana Y, Asif M, et al (2021) β-Caryophyllene induces apoptosis and inhibits angiogenesis in colorectal cancer models. Int J Mol Sci 22:. https://doi.org/10.3390/ijms221910550

Gayathri M, Madhan Shankar SR (2012) Anti-hemorrhoid, anti-inflammatory and anti-cancer activity of standardized leaf extract of Eupatorium triplinerve Vahl

Grabherr MG, Haas BJ, Yassour M, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883

Gupta D, Charles R, Garg SN (2004) Chemical composition of the essential oil from the leaves of eupatorium triplinerve vahl). Journal of Essential Oil Research 16:473–475. https://doi.org/10.1080/10412905.2004.9698774

Han JY, In JG, Kwon YS, Choi YE (2010) Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in Panax ginseng. Phytochemistry 71:36–46. https://doi.org/10.1016/j.phytochem.2009.09.031

Hilgers F, Habash SS, Loeschcke A, et al (2021) Heterologous production of ββ-caryophyllene and evaluation of its activity against plant pathogenic fungi. Microorganisms 9:1–19. https://doi.org/10.3390/microorganisms9010168

Hsu MJ, Peng SF, Chueh FS, et al (2019) Lupeol suppresses migration and invasion via p38/MAPK and PI3K/Akt signaling pathways in human osteosarcoma U-2 OS cells. Biosci Biotechnol Biochem 83:1729–1739. https://doi.org/10.1080/09168451.2019.1606693

Huang K, Jahani M, Gouzy J, et al (2023) The genomics of linkage drag in sunflower. Proceedings of the National Academy of Sciences 120:. https://doi.org/10.1101/2022.06.07.495047

Irrera N, D'ascola A, Pallio G, et al (2020) β-caryophyllene inhibits cell proliferation through a direct modulation of CB2 receptors in glioblastoma cells. Cancers (Basel) 12:. https://doi.org/10.3390/cancers12041038

Kim MJ, Jin J, Zheng J, et al (2015) Comparative transcriptomics unravel biochemical specialization of leaf tissues of Stevia for diterpenoid production. Plant Physiol 169:2462–2480. https://doi.org/10.1104/pp.15.01353

Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Mala, J. G. S., Zoghbi, M. G. B., Da Silva, M. H. L., & Andrade, E. H. A. (1999). Essential Oils of Eupatorium triplinerve Vahl and E. paniculatum Poepp. et Endl. Journal of Essential Oil Research, 11(5), 541-544:. https://doi.org/10.1080/10412905.1999.9701210

Mannino F, Pallio G, Corsaro R, et al (2021) Beta-caryophyllene exhibits anti-proliferative effects through apoptosis induction and cell cycle modulation in multiple myeloma cells. Cancers (Basel) 13:. https://doi.org/10.3390/cancers13225741

Maurya SK, Fatma H, Maurya AK, et al (2022) Role of lupeol in chemosensitizing therapy-resistant prostate cancer cells by targeting MYC, β-catenin and c-FLIP: in silico and in vitro studies. In Silico Pharmacol 10:. https://doi.org/10.1007/s40203-022-00131-3

Muthusamy S, Vetukuri RR, Lundgren A, et al (2020) Transient expression and purification of β-caryophyllene synthase in Nicotiana benthamiana to produce β-caryophyllene in vitro. PeerJ 2020:. https://doi.org/10.7717/peerj.8904

Phillips MA, Walter MH, Ralph SG, et al (2007) Functional identification and differential expression of 1-deoxy-D-xylulose 5-phosphate synthase in induced terpenoid resin formation of Norway spruce (Picea abies). Plant Mol Biol 65:243–257. https://doi.org/10.1007/s11103-007-9212-5

Prasad N, Sabarwal A, Yadav UCS, Singh RP (2018) Lupeol induces S-phase arrest and mitochondria-mediated apoptosis in cervical cancer cells. J Biosci 43:249–261. https://doi.org/10.1007/s12038-018-9743-8

Reinsvold RE, Jinkerson RE, Radakovits R, et al (2011) The production of the sesquiterpene β-caryophyllene in a transgenic strain of the cyanobacterium Synechocystis. J Plant Physiol 168:848–852. https://doi.org/10.1016/j.jplph.2010.11.006

Suja. G, Vinoth Kumar. T, Suganya. P (2019) Evaluation of Phytochemicals and In Vitro Anticancer Activity of Eupatorium Triplinerve. International Journal of Scientific Research in Biological Sciences 6:15–18. https://doi.org/10.26438/ijsrbs/v6i5.1518

Vranová E, Coman D, Gruissem W (2013) Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol 64:665–700

Walter MH, Hans J, Strack D (2002) Two distantly related genes encoding 1-deoxy-D-xylulose 5-phosphate synthases: Differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. Plant Journal 31:243–254. https://doi.org/10.1046/j.1365-313X.2002.01352.x

Wu J, Xu R, Lu J, et al (2022) Molecular cloning and functional characterization of two squalene synthase genes in Atractylodes lancea. Planta 255:. https://doi.org/10.1007/s00425-021-03797-9

Yang CH, Huang YC, Tsai ML, et al (2015) Inhibition of melanogenesis by β-caryophyllene from lime mint essential oil in mouse B16 melanoma cells. Int J Cosmet Sci 37:550–554. https://doi.org/10.1111/ics.12224

Zhang X, Gao Z, Chen K, et al (2022) Lupeol inhibits the proliferation and migration of MDA-MB-231 breast cancer cells via a novel crosstalk mechanism between autophagy and the EMT. Food Funct. https://doi.org/10.1039/d2fo00483f

306

307 **Figure 1-** Gene expression analysis of selected genes from A. triplinervis EF1: Elongation
308 factor 1 alpha, CS1: Caryophyllene synthase 1, LS: Lupeol synthase, FPPS1: Farnesyl
309 pyrophosphate synthase 2, FPPS2: Farnesyl pyrophosphate synthase 2 SQS1: Squalene
310 synthase 2, SQS2: Squalene synthase 2, SE1: Squalene epoxidase 1, SE2: Squalene epoxidase
311 2