

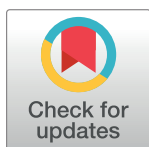
RESEARCH ARTICLE

Comparative transcriptome analysis to identify putative genes involved in carvacrol biosynthesis pathway in two species of *Satureja*, endemic medicinal herbs of Iran

Somayeh Shams¹, Ahmad Ismaili^{1*}, Farhad Nazarian Firouzabadi¹, Hasan Mumivand², Karim Sorkheh³

1 Faculty of Agriculture, Department of Plant Production and Genetic Engineering, Lorestan University, Khorramabad, Iran, **2** Faculty of Agriculture, Department of Horticultural Science, Lorestan University, Khorramabad, Iran, **3** Faculty of Agriculture, Department of Plant Production and Genetic Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

* ismaili.a@lu.ac.ir



OPEN ACCESS

Citation: Shams S, Ismaili A, Firouzabadi FN, Mumivand H, Sorkheh K (2023) Comparative transcriptome analysis to identify putative genes involved in carvacrol biosynthesis pathway in two species of *Satureja*, endemic medicinal herbs of Iran. PLoS ONE 18(7): e0281351. <https://doi.org/10.1371/journal.pone.0281351>

Editor: Mounir Tilouai, Waterford Institute of Technology, IRELAND

Received: July 7, 2022

Accepted: January 22, 2023

Published: July 7, 2023

Copyright: © 2023 Shams et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data that support the findings of this study are openly available in PRJNA464236, PRJNA464241 at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA464236/> and <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA464241>.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Satureja is rich in phenolic monoterpenoids, mainly carvacrol, that is of interest due to diverse biological activities including antifungal and antibacterial. However, limited information is available regarding the molecular mechanisms underlying carvacrol biosynthesis and its regulation for this wonderful medicinal herb. To identify the putative genes involved in carvacrol and other monoterpene biosynthesis pathway, we generated a reference transcriptome in two endemic *Satureja* species of Iran, containing different yields (*Satureja khuzistanica* and *Satureja rechingeri*). Cross-species differential expression analysis was conducted between two species of *Satureja*. 210 and 186 transcripts related to terpenoid backbone biosynthesis were identified for *S. khuzistanica* and *S. rechingeri*, respectively. 29 differentially expressed genes (DEGs) involved in terpenoid biosynthesis were identified, and these DEGs were significantly enriched in monoterpenoid biosynthesis, diterpenoid biosynthesis, sesquiterpenoid and triterpenoid biosynthesis, carotenoid biosynthesis and ubiquinone and other terpenoid-quinone biosynthesis pathways. Expression patterns of *S. khuzistanica* and *S. rechingeri* transcripts involved in the terpenoid biosynthetic pathway were evaluated. In addition, we identified 19 differentially expressed transcription factors (such as MYC4, bHLH, and ARF18) that may control terpenoid biosynthesis. We confirmed the altered expression levels of DEGs that encode carvacrol biosynthetic enzymes using quantitative real-time PCR (qRT-PCR). This study is the first report on *de novo* assembly and transcriptome data analysis in *Satureja* which could be useful for an understanding of the main constituents of *Satureja* essential oil and future research in this genus.

Introduction

Terpenoids are the most abundant compounds of plant secondary metabolites and have been interesting due to their pharmacological utility [1]. The universal precursors of terpenoids

consist of IPP (isopentenyl pyrophosphate) and its isomer, DMAPP (dimethylallyl pyrophosphate) [2]. The IPP precursor in plants is produced by two completely separate pathways; mevalonic acid (MVA) and the methylerythritol phosphate (MEP) pathways [3]. The condensation of DMAPP and IPP provides geranyl diphosphate (GPP). The GPP is the immediate precursor to monoterpenes, diterpenes, and tetraterpenoids. They are the main components of many essential oils extensively used as fragrances, cosmetics ingredients, food additives, and the synthesis of perfume chemicals [4].

Carvacrol is a monoterpene compound derived from isoprene hydrocarbure (2-methyl-1, 3-butadiene) and formed by the attachment of two or more isoprene molecules [5]. Carvacrol (over 90%) is the major essential oil produced by *Satureja* species [6]. It is produced in the essential oils of some members of Lamiaceae family such as Marjoram [7], Fern leaf lavender [8] and Thymes [9].

Satureja (savory) is an important genus belonging to the Lamiaceae family [10], it consists of more than 200 species, including nine species of endemics to Iran. *Satureja khuzistanica* is an aromatic endemic species. This species was first illustrated in 1944 as a new species in the flora of Iran [11]. *S. khuzistanica*, as a medicinal herb, is distributed in mountainous slopes in north, northwest, northeast, central and southern parts of Iran [12]. *S. rechingeri* is another endemic where it is found only in Ilam province, west of Iran [6].

In extensive studies on *Satureja* species, the biological activities [13], including their hypotensive and hypolipidemic [14] effects, antifungal [15], anti-bacterial [16], anti-oxidant [17], anti-parasitic [17], and anti-inflammatory activities [18] have been demonstrated. The importance medicinal of savory has been attributed to the accumulation of carvacrol in essential oil and free phenolic acids, especially rosmarinic acid [19].

Previous studies have focused on the phytochemical analysis [20, 21] and pharmacological activities of savory essential oil [13] while genes involved in the biosynthesis pathway of essential oil compounds are poorly investigated in this plant.

RNA sequencing (RNA-Seq) is a high throughput technique based on next-generation sequencing (NGS) technology [22]. mRNA sequencing can provide transcripts information to examine gene expression levels and can help us to understand more about functional genes in the transcriptome study of organisms without the genomic sequence [23]. RNA-Seq strategy can be used to identify and characterize secondary metabolite genes and their pathways [24, 25]. Transcriptome analysis using RNA-Seq had been applied in medicinal herbs such as *Salvia miltiorrhiza* [26], *Dracocephalum tanguticum* [27], *Cymbopogon flexuosus* [28], *Hovenia acerba* [29], *Swertia mussotii* [30] and *Piper nigrum* [31].

Despite its medicinal value, knowledge about the genetics of *Satureja* is limited, which has hindered the investigation of metabolite synthesis. Therefore, comprehensive research is important to increase the biological understanding of this genus and to clarify the metabolic pathways leading to the biosynthesis of valuable compounds [30].

In the present study, *S. khuzistanica* and *S. rechingeri* were used as the experimental materials to establish a representative model of transcriptome sequences of *Satureja* using Illumina sequencing. The transcriptome database was used as reference data to identify candidate genes associated with carvacrol biosynthesis. Based on this, we used differential expression profile analysis to compare the transcripts involved in carvacrol biosynthesis. The transcriptome sequences and gene expression profiles provide a database for functional genomic studies on *Satureja* in the future and will facilitate a better understanding of the molecular mechanisms of secondary metabolites biosynthetic pathways.

Materials and methods

Plant material

S. khuzistanica and *S. rechingeri* were collected from Khorramabad city, Lorestan province, Iran. In the present study, leaf tissue (the main part of the biosynthesis of secondary metabolites) was used to generate the transcriptome of *Satureja*. The leaves were collected during the full flowering stage (The phase containing the highest essential oil content) [32–34]. Three biological replicates were considered for each species. All samples were snap-frozen in liquid nitrogen and stored at -80°C until used.

RNA extraction

Total RNA was isolated from all replicates of each species with TRIzol Reagent (Invitrogen, Carlsbad, California, United States), according to the manufacturer's protocol, and then DNase I (Takara, Japan) was used to eliminate genomic DNA. The RNA quantity and quality were determined with Agilent 2100 Bioanalyzer (Agilent Technologies, USA). A portion of the extracted RNA was sequenced in each biological replication, and the remaining part was used for qRT-PCR analysis.

RNA-Seq library preparation and Illumina sequencing

The libraries were created with high-quality total RNA samples ($\text{OD}_{260/280} = 1.8 \sim 2.2$; $\text{OD}_{260/230} \geq 2.0$; RNA Integrity Number (RIN) ≥ 8 out of 10; $28\text{S}:18\text{S} > 1.0$). The cDNA libraries were prepared according to Illumina's protocol (Illumina Inc., San Diego, USA; cat. no. RS-100-0801) by a genome sequencing company (Novo gene, China) and were sequenced to a depth of 40 million paired-end reads per each biological replicate using an Illumina HiSeq™ 2500 platform (150 bp; Illumina).

The transcriptome raw sequencing data of *S. khuzistanica* and *S. rechingeri* have been submitted to the NCBI GenBank (<https://www.ncbi.nlm.nih.gov>) with BioSample accession numbers as following: SAMN09079833, SAMN09079834, SAMN09079835 and SAMN09080197, SAMN09080198, SAMN09080199 and BioProject accession numbers: PRJNA464236, PRJNA464241.

RNA-Seq data processing and *de novo* assembly

The raw read quality was evaluated using the FastQC software (version 0.11.5) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) [35]. High-quality reads were filtered by removing the adaptor and low-quality sequences (Q score < 20) from the raw reads using the Trimmomatic tool (v. 0.36) [36]. High-quality reads were assembled using Trinity *de novo* assembler (v2.8.5) (<http://trinityrnaseq.sourceforge.net/>) [37] with default parameters and the k-mer size of 25. The result of assembled sequences as transcripts were filtered through trinity to have FPKM (Fragments per kilobase of transcript sequence per millions base pairs sequenced) more than one value per gene. To validate and assess the *de novo* assembly quality, the reads were mapped to the assembled transcripts using Bowtie 2–2.3.4.1 package with the default parameters [38].

Functional annotation of assembled transcripts

The Trinotate pipeline (<https://trinotate.github.io/>) was used for gene functional annotation. All of the assembled transcripts were searched against several databases, including NCBI Non-redundant (NR) (<http://www.ncbi.nlm.nih.gov/>) and Swiss-Prot [39] through BLASTx with an e-value cut-off of $1e^{-5}$ and only the best homologue was reported. Open reading frames (ORFs) were predicted using TransDecoder (v5.5.0) (<http://transdecoder.sourceforge.net/>)

following the identification of the longest ORFs. Protein sequences of predicted ORFs were searched using BLASTp against Swiss-Prot and eggNOG database with an e-value cut-off of $1e^{-3}$, seed_score = 80, and query_cov = 60 [40]. Online KEGG Automatic Annotation Server (v2.1) (KASS) (https://www.genome.jp/kaas-bin/kaas_main?mode=partial) with P-value cut-off of ≤ 0.05 was performed to determine Pathways, KO assignments, and functional descriptions of assembled transcripts [41]. Protein domains were identified by searching against Pfam-A HMMs database with default parameters and e-value cut-off of $1e^{-3}$ [42].

Identification of differentially expressed genes (DEGs)

To evaluate the abundance of transcripts in each species, transcripts quantification for each biological replicate was performed by RSEM software [43] and bowtie2 was used for alignment, and then the gene expression matrix was built to generate a normalized expression value. The ExN50 analysis was used to estimate the expression levels of the transcripts. DEGs were identified from the count's matrix estimated by Integrated Differential Expression Analysis MultiEXperiment (IDEAmax) web server (<http://zazil.ibt.unam.mx/idea>) with four methods: edgeR, NOISeq, limma, and DESeq2 [44]. The DEGs with log 2-fold change ≥ 2 (up-regulated genes) and ≤ -2 (down-regulated genes) and P-value cut-off of ≤ 0.01 were screened, and only transcripts that were differentially expressed according to the four methods considered as DEGs. Venny (v2.0) was used to compare the four methods results of DEGs [45]. Web Gene Ontology Annotation Plot (WEGO) was used for Gene Ontology (GO) terms [46] and g:Profiler was used to perform GO and KEGG enrichment analysis (Organism: *Arabidopsis thaliana*, Statistical domain scope: only annotated genes, Significance threshold: g:SCS algorithm, P-value cut-off of < 0.01) [47].

Identification of genes related to essential oil compounds biosynthesis

Candidate transcripts of secondary metabolites in *S. khuzistanica* and *S. rechingeri* were identified using BLASTx (length > 80 bp, percentage of identical matches > 70 mismatch < 150 and, e-value cut-off of $1e^{-5}$) against NCBI NR and Swiss-Prot databases and BLASTp (Assignment method: SBH (single-directional best hit), manually annotated organisms: Monocots and Eudicots, P-value cut-off of ≤ 0.05) against KEGG Automatic Annotation Server Ver. 2.1. The Trinotate pipeline was used to assign GO terms to the annotated sequences to predict the functions of the sequences. All of the transcripts related to terpenoid biosynthesis pathways were identified based on the annotation results. Expression values of transcripts involved in the terpenoid backbone biosynthesis and monoterpenoid biosynthesis pathways of *Satureja* were illustrated as a heatmap using HemI software (Heatmap Illustrator (v1.0)) [48]. Enriched GO-terms were visualized with REVIGO with default parameters [49]. All of the assembled transcripts were searched using BLASTx (e-value cut-off of $1e^{-5}$) against Plant TF Database v5 (PlantTFDB) [50]. TFs were filtered by alignment length > 80 bp, percentage of identical matches > 70 , e-value cut-off of $1e^{-5}$, number of mismatches < 150 bp and pct_hit_len_aligned > 50 . The differentially expressed genes associated with terpene biosynthesis was shown as a heat map using the Clustergrammer (Distance: correlation, Linkage type: complete) (<https://maayanlab.cloud/clustergrammer/>) [51]. Co-expression interactions networks were constructed using the GeneMANIA prediction web server, with default parameters (Organism: *Arabidopsis thaliana*) (<https://genemania.org/>) [52].

Quantitative real-time PCR (qRT-PCR) validation

The qRT-PCR technique was used for expression analysis of selected genes of *Satureja* using the real-time PCR detection system (Step One Real-Time PCR System, Applied Biosystems,

USA) and TaKaRa SYBR® Green Permixon Taq™ II. Primers were designed using Primer premier6 software. Specific oligonucleotides for qRT-PCR analysis are listed in the [S1 Table](#). The thermal profile was 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. Elongation factor1a (EF1a) as the reference gene was used for qRT-PCR normalization [53]. Data were analyzed using the comparative cycle threshold (Ct) method [54, 55]. Two technical replicates were used for each biological replicate. Statistical analyses were performed using ANOVA with IBMSPSS 24.0 (SPSS, Inc., Chicago, IL, United States). A significance level of P-value cut-off of < 0.05 was applied.

Essential oil extraction

Fresh leaf tissue (200 g) of *S. khuzistanica* and *S. rechingeri* was air-dried for 7 days. The essential oil of these plants was extracted using Clevenger for 4 hours. The extracted essential oil was dehydrated over Na₂SO₄ and then stored in tightly closed, dark vials at 4°C until analysis [56].

GC-MS analysis of essential oil. The GC-MS (Gas chromatography-mass spectrometry) analysis was performed with a Thermoquest Finnigan Trace. A fused silica capillary DB-5 column with 30 m long and 0.25 mm diameter was used. The column temperature reached 60°C to 250°C. Nitrogen carrier gas flowed at a rate of 1 ml/min and the transfer-line temperature was 250°C [56].

The identification of essential oil components was performed by comparing their mass with those of the internal reference mass spectra library and confirmed by comparison of their retention indices (RI) with those reported in the literature. Co-injection of essential oils with authentic standards (Sigma-Aldrich, USA) was performed. Quantification of the relative percentage of essential oil constituents was performed by GC-FID peak area normalization considering an equal response factor for the different chemical classes occurring in the essential oils [57]. The Phytochemical analysis was performed in three replicate for each species. Differences were considered statistically significant when the P-value cut-off of < 0.05. Statistical analysis was performed using IBMSPSS 24.0 (SPSS, Inc., Chicago, IL, United States).

Results

RNA-Seq library preparation and sequencing

Short-read transcriptome sequencing from leaf tissue of *S. khuzistanica* and *S. rechingeri* was carried out ([Fig 1](#)). Over 130 million of 150 bp paired-end reads were acquired for each species. After filtering and deleting poor-quality reads, more than 95% of paired-end reads were considered for downstream analysis ([Table 1](#)).

De novo assembly. In the present study, two strategies were considered for *de novo* transcriptome assembly. The first strategy involved one assembled transcript for each biological replicate. For the second strategy, to obtain a more complete assembly, the reads of three replicates of each species were merged (one assembly per species) ([S2 Table](#)). The combined assemblies showed higher N50 values than transcripts assembled for each biological replicate. Therefore, the study was performed with data from the second strategy. The transcript assembly details are presented in [Table 2](#). After assembling the reads, the average length of the transcripts in *S. khuzistanica* and *S. rechingeri* leaf transcriptome was 1079 bp (N50 = 1802 bp) and 1018 bp (N50 = 1740 bp), and the average length of them based on only longest isoform per 'gene' was 819 bp (N50 = 1504 bp) and 773 bp (N50 = 1426 bp), respectively ([Table 2](#)). According to the mapping results, 98.48% and 98.41% of reads were correctly mapped to the assembled transcripts of *S. khuzistanica* and *S. rechingeri*, respectively.

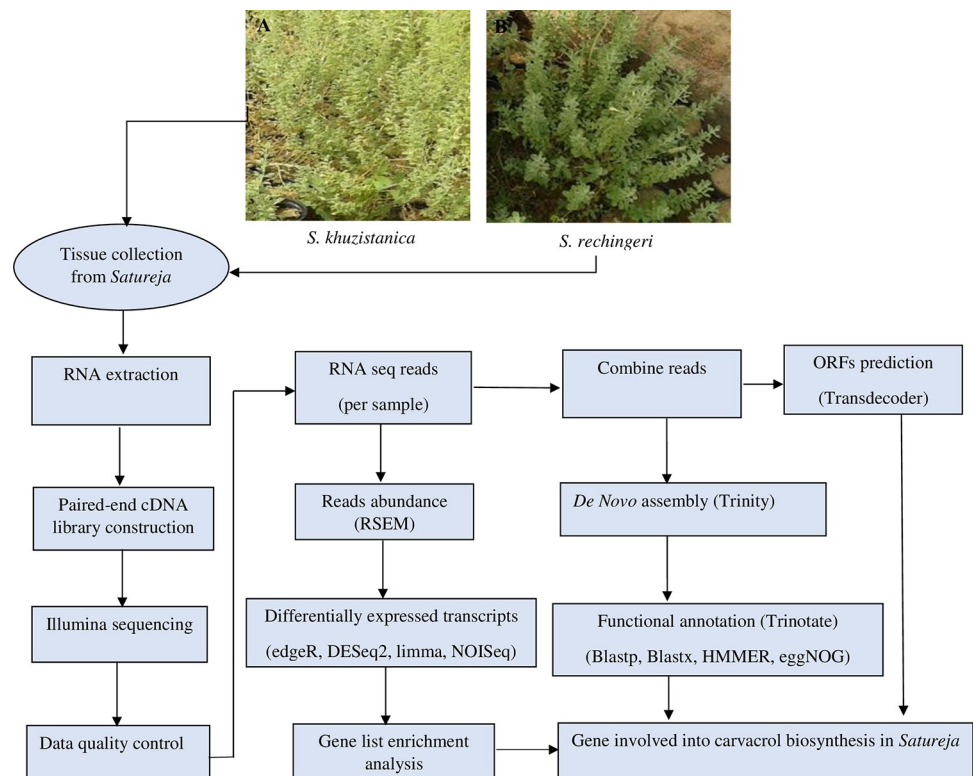


Fig 1. Plant materials. A) *S. khuzistanica*, B) *S. rechingeri*, Schematic of *de novo* RNA-seq analysis workflow.

<https://doi.org/10.1371/journal.pone.0281351.g001>

Homology search and functional annotation

The annotation was performed using BLASTx against Swiss-Prot and NCBI NR databases (S3 Table). The results showed that 106255 (59.98%) transcripts for *S. Khuzistanica* and 102908 (67.11%) transcripts for *S. rechingeri* have at least one significant hit against NCBI Nr database. The number of transcripts with open reading frames in each species was identified using the TransDecoder pipeline (S1 Fig). The top 10 most frequent domain families of *S. khuzistanica* and *S. rechingeri* were protein kinase followed by PPR_2, Pkinase_Tyr, NB-ARC, p450, LRR_8, RRM_1, WD40, Myb_DNA-binding, and PPR (S2 Fig). The global and overview KEGG pathway maps that detected in *S. khuzistanica* and *S. rechingeri* is shown in Table 3.

Identification of differentially expressed genes (DEGs)

To reduce the number of comparisons, 9125 genes were differentially expressed according to the four statistical methods (edgeR, NOISeq, limma, and DESeq2) and were considered DEGs

Table 1. Number of paired-end raw reads and trim summary per each biological replicate.

Species	Specie ID of each replicate	Raw reads (PE)	Trimmed reads (PE)	Percentage trimmed (%)
<i>Saturejakhuzistanica</i>	Sk_1	43646790	41761392	95.68
	Sk_2	43057192	41444554	96.25
	Sk_3	43881896	42512350	96.88
<i>Saturejarechingeri</i>	Sr_1	44198084	42597116	96.38
	Sr_2	44659238	43229780	96.80
	Sr_3	45212806	43923756	97.15

<https://doi.org/10.1371/journal.pone.0281351.t001>

Table 2. Trinity assembly statistics report of combined assembled transcripts.

Counts of transcripts, etc.					
<i>Satureja khuzistanica</i>			<i>Satureja rechingeri</i>		
Total trinity 'genes'	Total trinity 'transcripts'	Percent GC	Total trinity 'genes'	Total trinity 'transcripts'	Percent GC
121192	271907	43.93	119203	225534	45.02
Trinity assembly stats					
Stats	based on all transcript contigs	based on only LONGEST ISOFORM per 'GENE'	based on all transcript contigs	based on only LONGEST ISOFORM per 'GENE'	
Contig N10	4274	4032	4225	3910	
Contig N20	3241	2980	3173	2868	
Contig N30	2628	2345	2556	2241	
Contig N40	2173	1893	2108	1795	
Contig N50	1802	1504	1740	1426	
Median contig length	652	418	584	385	
Average contig length	1079.12	818.89	1018.04	772.92	
Total assembled bases	293419043	99242431	229602486	92134478	

<https://doi.org/10.1371/journal.pone.0281351.t002>

(Fig 2). Out of 9125 differentially expressed genes between the two species, 7026 transcripts were annotated against Swiss-Prot database (S4 Table). Among these DEGs, 4133 genes were up-regulated and 2893 genes were down-regulated under *S. khuzistanica* vs *S. rechingeri*.

GO enrichment analysis for differentially expressed genes

In total, the 4594 DEGs were assigned to three Gene ontology (GO) terms (biological process, molecular functions, and cellular component) (Fig 3). Most of the DEGs were classified in binding (GO: 0005488) and catalytic activity (GO: 0003824) in the molecular function category. In the biological process category, the majority of the GO terms were related to the metabolic process (GO: 0008152) followed by cellular process (GO:0009987). For the cellular

Table 3. Global and overview KEGG pathway maps of *S. khuzistanica* and *S. rechingeri* transcripts defined by KAAS.

KEGG Pathways	(Ko ^a)	NGKP ^b	NIGKPSK ^c	NIGKPSR ^d
Metabolic pathways	01100	4121	935	936
Biosynthesis of secondary metabolites	01110	1179	425	425
Microbial metabolism in diverse environments	01120	1180	155	154
Biosynthesis of antibiotics	01130	1016	203	203
Carbon metabolism	01200	353	98	98
2-Oxocarboxylic acid metabolism	01210	82	29	29
Fatty acid metabolism	01212	84	29	31
Biosynthesis of amino acids	01230	232	100	100
Degradation of aromatic compounds	01220	215	5	6
Total	-	8247	1979	1982

^aKo: Reference KEGG pathway number

^bNGKP: Number of genes in each KEGG pathway

^cNIGKPSK: Number of identified genes in KEGG pathway of *S. khuzistanica*

^dNIGKPSR: Number of identified genes in KEGG Pathway of *S. rechingeri*

<https://doi.org/10.1371/journal.pone.0281351.t003>

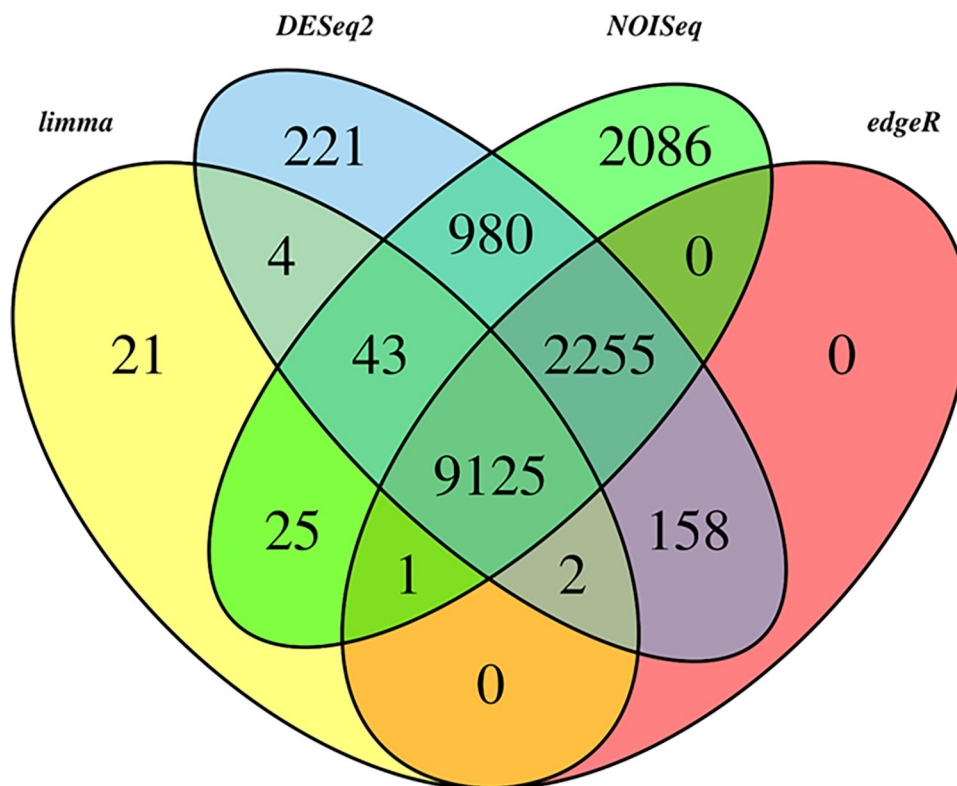


Fig 2. Venn diagram of four statistical methods (edgeR, NOISeq, limma, and DESeq2) of differentially expressed genes in *S. khuzistanica* vs *S. rechingeri*.

<https://doi.org/10.1371/journal.pone.0281351.g002>

components, the majority of the GO terms were related to membrane (GO: 0016020), and cell part (GO: 0044464) (Fig 3).

GO enrichment analysis was done on the DEGs. In all, 1318 GO terms were used for enrichment from two *Satureja* species (S5 Table). In the biological process class, the enriched categories were related to cellular process (GO:0009987), metabolic processes (GO:0008152), organic substance metabolic process (GO:0071704), developmental process (GO:0032502), cellular metabolic process (GO:0044237) and others (S5 Table). The presence of GO categories related to secondary metabolite biosynthetic process (GO:0044550), isopentenyl diphosphate biosynthetic process, mevalonate pathway (GO:0019287), and Acetyl-CoA metabolic process (GO:0006084) led us to identify the genes related to terpenoid biosynthesis in *Satureja* species (S5 Table). GO enrichment was analyzed for DEGs related to the terpenoid biosynthesis in *Satureja* (S6 Table). As a complementary approach to the GO analysis, the scatterplot of the selected terpenoid biosynthesis view of GO category enrichment analysis was constructed based on DEGs related to the biological process of *S. khuzistanica* vs *S. rechingeri* (Fig 4). the most enriched biological process terms for over-presented DEGs were cellular lipid metabolic process (GO:0044255), organic cyclic compound biosynthetic process (GO:1901362), and cellular aromatic compound metabolic process (GO:0006725). Among all the sets in the biological process, three enriched categories were related to the monoterpene metabolism, GO:0043693 as a monoterpene biosynthetic process, GO:0043692 as a monoterpene metabolic process, and GO:1902767 as an isoprenoid biosynthetic process via mevalonate in *S. khuzistanica* vs *S. rechingeri* (S6 Table).

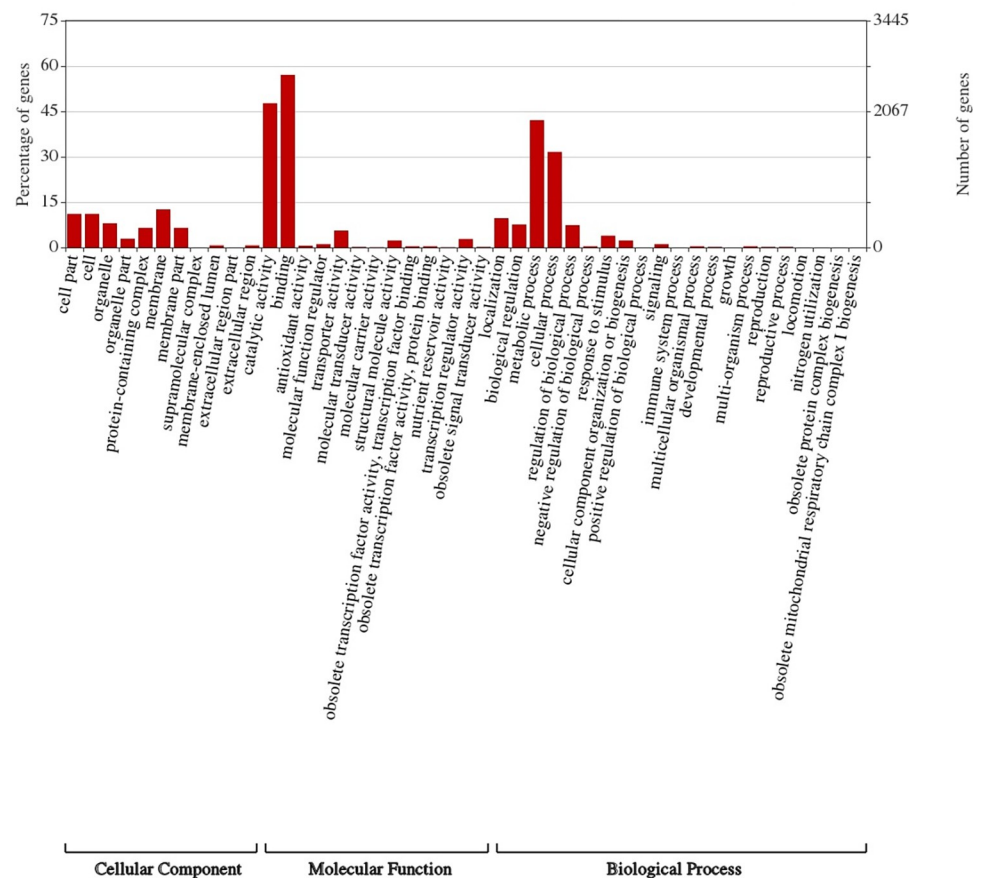


Fig 3. Gene ontology (GO) classification of annotated DEGs. The x-axis indicates the subgroups in GO terms; y-axis shows the percentages of genes (number of a particular gene divided by total gene number).

<https://doi.org/10.1371/journal.pone.0281351.g003>

Pathway enrichment analysis for differentially expressed genes

The protein sequence of predicted ORFs of *S. khuzistanica* (91467 transcripts) and *S. rechingeri* (78733 transcripts) were searched to identify transcripts related to secondary metabolic pathways. The results showed that 44.29% of all transcripts were assigned to 401 KEGG pathways.

To identify the biological pathways, DEGs were mapped to KEGG pathways. In all, 4140 DEGs were assigned to 374 KEGG pathways (S7 Table). The highest number of DEGs was assigned to metabolic pathways followed by biosynthesis of secondary metabolites, microbial metabolism in diverse environments, and carbon metabolism (Fig 5). The most common enriched pathways were ‘metabolites biosynthesis’ and ‘biosynthesis of secondary metabolites’.

Pathway enrichment analysis for DEGs related to the terpenoid pathway in *Satureja* was used to identify significant KEGG metabolic pathways (Tables 4 and 5). Based on the results of the pathway enrichment analysis, 74 terpenoid biosynthesis-related transcripts were identified (S8 Table). Of KEGG secondary metabolic pathways, most transcripts were assigned to “phenylpropanoid biosynthesis” (ko00940), “terpenoid backbone biosynthesis” (ko00900), and “Flavonoid biosynthesis” (ko00941) in *S. khuzistanica* and *S. rechingeri* leaf transcriptome (Table 5 and S9 Table).

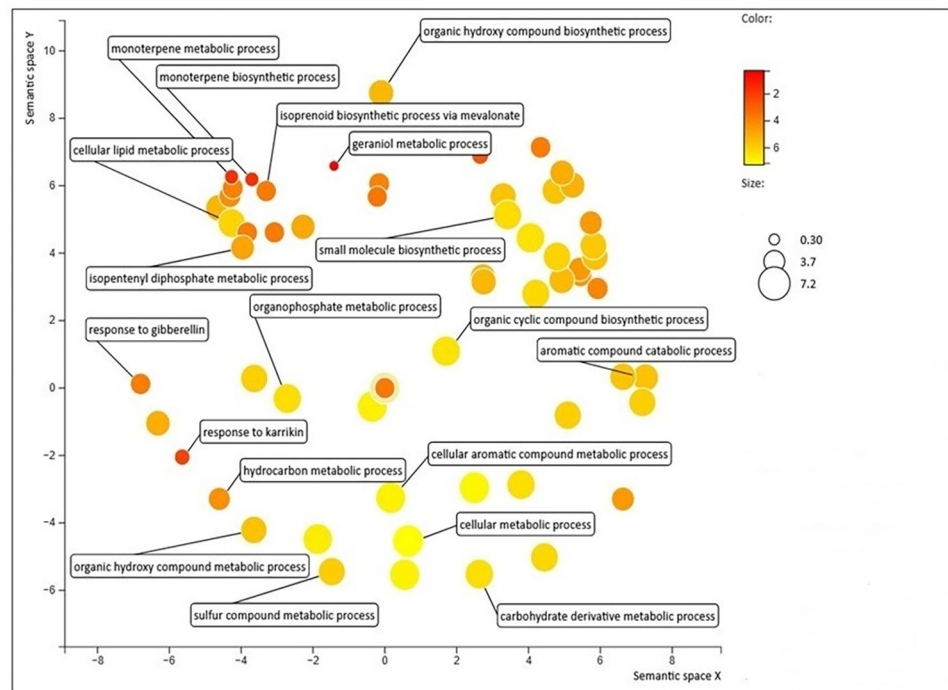


Fig 4. The scatterplot of selected terpenoid biosynthesis view of GO category enrichment analysis of differentially expressed genes related to biological process found in *S. khuzistanica* vs *S. rechingeri* produced by REVIGO.

Circles depicted by filled color show significantly enriched GO terms with $\log_{10}$ p-value < 0.05 . The color of the bubbles shows the P-value and the size of the bubbles shows the frequency of the GO term in the underlying GOA database.

<https://doi.org/10.1371/journal.pone.0281351.g004>

Identification of genes encoding key enzymes involved in the monoterpenoid biosynthesis pathway

In this study, 210 and 186 transcripts isoforms related to terpenoid backbone biosynthesis were identified for *S. khuzistanica* and *S. rechingeri*, respectively, which have not been identified in this genus (S10 Table). Transcriptomic analysis to investigate monoterpenoid accumulation patterns in two species of *Satureja* (*S. khuzistanica* and *S. rechingeri*) discovered 37 transcripts in the mevalonate (MVA) pathway, including nine key enzymes, and 35 transcripts in the methyl erythritol 4-phosphate (MEP) pathway, involving seven key enzymes. It inferred more differences between different species from the same genus.

In the MVA pathway, 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) as the key enzyme of this pathway contained the highest number of transcripts (*Sk*: 14 transcripts and *Sr*: 9 transcripts) followed by 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) (*Sk*: 10 transcripts and *Sr*: 7 transcripts).

In the MEP pathway, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR) (9 transcripts), followed by 1-deoxy-D-xylulose-5-phosphate synthase (DXS) (*Sk*: 8 transcripts and *Sr*: 7 transcripts) as a catalyst enzyme initiates the MEP pathway contained the highest number of transcripts [58] whereas the number of transcripts for 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (MCT), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK), and 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS) was lower (S10 Table).

We performed differential expression analysis using transcriptome data, to identify candidate DEGs encoding key enzymes in the monoterpenoid biosynthesis pathways. DEGs

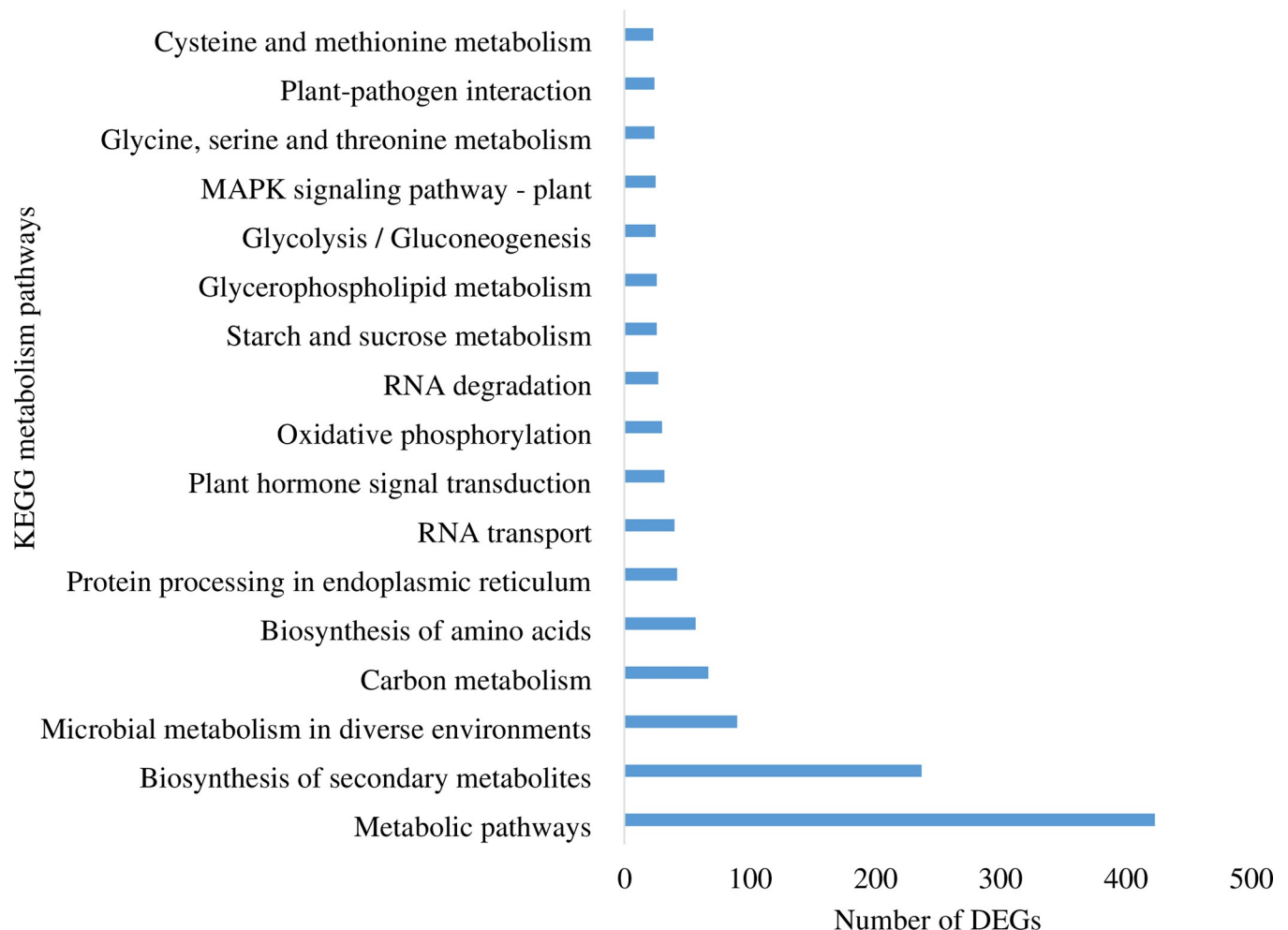


Fig 5. Top hits of KEGG metabolism pathway categories of DEGs.

<https://doi.org/10.1371/journal.pone.0281351.g005>

involved in “Biosynthesis of secondary metabolites” were overexpressed in pair comparisons; DEGs related to terpenoid biosynthetic pathways were detected in *S. khuzistanica* vs *S. rechingeri* (Fig 6). The 28 key genes related to monoterpenoid biosynthesis, diterpenoid biosynthesis, sesquiterpenoid and triterpenoid biosynthesis, carotenoid biosynthesis, and Ubiquinone and other terpenoid-quinone biosynthesis were discovered by DEGs analysis between *S. khuzistanica* vs *S. rechingeri*.

As is shown in Fig 6, out of seven genes involved in the MEP biosynthesis pathway, four genes have the highest upregulation in *S. rechingeri*. The expression of the DXS gene in *S. rechingeri* was double of *S. khuzistanica*. The expression of both (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (HDS) and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (HDR) genes in *S. rechingeri* was markedly higher *S. rechingeri* compared to that in *S. khuzistanica*. The expression of the CMK gene in *S. rechingeri* was double of *S. khuzistanica*. The expression value of MCT and acetyl-CoA C-acetyltransferase (ACCT) genes was almost equal in both *S. khuzistanica* and *S. rechingeri*. In the MVA pathway, the expression of the HMGS gene in *S. khuzistanica* is higher than in *S. rechingeri*, like the HMGR gene. Interestingly, HMGS and phosphomevalonate kinase (PMK) exhibited the highest expression level in *S. khuzistanica*.

Table 4. KEGG pathway enrichment analysis of differentially expressed genes.

Enriched pathway term	KEGG ID	P-value	Term size ^a	Query size ^b
Biosynthesis of secondary metabolites	01110	8.81911E-10	1236	17
Ubiquinone and other terpenoid-quinone biosynthesis	00130	4.45951E-09	39	15
Terpenoid backbone biosynthesis	00900	4.45951E-09	61	17
Monoterpenoid biosynthesis	00902	0.000274219	7	9
Isoquinoline alkaloid biosynthesis	00950	0.000999409	22	6
Phenylalanine metabolism	00360	0.001892402	33	6
Tropane, piperidine and pyridine alkaloid biosynthesis	00960	0.001932288	36	6
Tyrosine metabolism	00350	0.002194744	41	6
Diterpenoid biosynthesis	00904	0.003544646	21	16
Phenylalanine, tyrosine and tryptophan biosynthesis	00400	0.003544646	56	6
Synthesis and degradation of ketone bodies	00072	0.005263202	4	4
Metabolic pathways	01100	0.006103447	2273	16
Cysteine and methionine metabolism	00270	0.011268698	120	6
Butanoate metabolism	00650	0.020577188	20	4
Biosynthesis of cofactors	01240	0.036374303	250	14
Valine, leucine and isoleucine degradation	00280	0.04636249	52	4

^aTerm size: all genes of *Arabidopsis thaliana* associated with the given term

^bQuery size: A set of genes as a query.

<https://doi.org/10.1371/journal.pone.0281351.t004>

Secondly, DEGs analyses show that the expression pattern of the core genes encoding enzymes involved in the monoterpene synthases required for carvacrol biosynthesis are expressed at high levels in *S. rechingeri* compared to *S. khuzistanica*. The expression level of farnesyl pyrophosphate synthase (FPPS) in *S. khuzistanica* was higher than that in *S. rechingeri*, which was consistent with the expression pattern of DEGs in the MVA pathway. The expression level of geranyl pyrophosphate synthase (GPPS) in *S. khuzistanica* was lower than in *S. rechingeri*.

Of the 28 key genes, the transcript level of terpene synthases (TPS), Cytochrome P450 71D18 (CYP71D18), and carveol dehydrogenase (DH) genes in the monoterpene biosynthesis pathway were significantly up-regulated in *S. rechingeri* vs *S. khuzistanica* (Fig 6). The sequence of *Satureja* TPS is more than 80% similar to the known TPS of other plants in the mint family, including TPS2 of *O. vulgare* [59] and *Thymus* [60]. CYP71D18 of *Satureja* is more than 80% similar to limonene-3-hydroxylase in *Thymus vulgaris* (S3A and S3B Fig) [60].

Some DEGs were putatively involved in the monoterpenoids biosynthesis pathway (Table 6). The 1279 transcripts were annotated as CYP450s, of which 84 transcripts were differentially expressed in leaf tissues of *Satureja*. The 179 transcripts of TPS gene candidates were identified based on Swiss-Prot database, while 11 transcripts were differentially expressed in the leaf tissue of *Satureja*. The 2309 transcripts that encode the DH genes family were annotated, of which 140 transcripts showed differential expression in leaf tissues of *Satureja*.

Analysis of putative transcription factor genes related to terpenoid biosynthesis

Among the DEGs, some important genes were observed such as transcription factors. 226 DEGs associated with TFs gene families in the leaf tissue of *Satureja* were identified (S11 Table) and classified into 41 TF gene families of 34 plant species (S4A and S4B Fig). The top 10 hits of TFs genes families (61.50%, 139 transcripts) were showed by bHLH, MYB_related, NAC, ERF, bZIP, ARF, C3H, HSF, C2H2 and FAR1 (S4A Fig). Among the 41 TF gene families,

Table 5. Transcripts related to secondary metabolites biosynthesis in *S. khuzistanica* and *S. rechingeri* leaf transcriptome.

Category	KEGG pathways	Ko ^a	NGKP ^b	NIGKPSK ^c	NITKPSK ^d	NIGKPSR ^e	NITKPSR ^f
Metabolism of terpenoids and polyketides	Terpenoid backbone biosynthesis	00900	51	31	231	31	210
	Monoterpenoid biosynthesis	00902	38	4	131	4	98
	Sesquiterpenoid and triterpenoid biosynthesis	00909	73	8	63	7	40
	Diterpenoid biosynthesis	00904	66	9	114	9	97
	Carotenoid biosynthesis	00906	49	20	151	20	110
	Brassinosteroid biosynthesis	00905	11	8	79	8	66
	Insect hormone biosynthesis	00908	10	1	37	1	32
	Zeatin biosynthesis	00981	18	7	152	7	102
	Limonene and pinene degradation	00903	12	2	39	2	37
	Geraniol degradation	00281	15	2	9	2	13
	Biosynthesis of ansamycins	01051	32	1	16	1	15
	Polyketide sugar unit biosynthesis	00523	58	1	2	1	3
	Biosynthesis of siderophore group nonribosomal peptides	01053	30	1	4	1	2
Biosynthesis of other secondary metabolites	Phenylpropanoid biosynthesis	00940	37	22	620	21	544
	Stilbenoid, diarylheptanoid and gingerol biosynthesis	00945	12	5	153	5	153
	Flavonoid biosynthesis	00941	27	14	208	14	206
	Flavone and flavonol biosynthesis	00944	15	3	14	3	13
	Anthocyanin biosynthesis	00942	16	3	8	2	4
	Isoflavonoid biosynthesis	00943	17	3	33	2	21
	Indole alkaloid biosynthesis	00901	30	1	12	1	14
	Isoquinoline alkaloid biosynthesis	00950	62	8	125	8	125
	Tropane, piperidine and pyridine alkaloid biosynthesis	00960	26	9	80	9	67
	Caffeine metabolism	00232	16	3	9	3	11
	Betalain biosynthesis	00965	7	3	21	3	22
	Glucosinolate biosynthesis	00966	20	3	29	3	23
	Monobactam biosynthesis	00261	28	6	30	6	29
	Streptomycin biosynthesis	00521	19	4	45	4	40
	Neomycin, kanamycin and gentamicin biosynthesis	00524	57	1	24	1	18
	Novobiocin biosynthesis	00401	30	2	22	2	23
	Phenazine biosynthesis	00405	16	2	15	2	8
	Prodigiosin biosynthesis	00333	20	3	31	3	34
	Aflatoxin biosynthesis	00254	13	1	3	1	4
	Biosynthesis of secondary metabolites other antibiotics	00998		1	5	1	4
	Biosynthesis of secondary metabolites unclassified	00999		3	59	3	56

^aKo, Reference KEGG pathway number^bNGKP, Number of genes in each KEGG pathway^cNIGKPSK, Number of identified genes in KEGG Pathway of *S. khuzistanica*^dNITKPSK, Number of identified transcripts in KEGG Pathway of *S. khuzistanica*^eNIGKPSR, Number of identified genes in KEGG pathway of *S. rechingeri*^fNITKPSR, Number of identified transcripts in KEGG pathway of *S. rechingeri*.<https://doi.org/10.1371/journal.pone.0281351.t005>

bHLH (11%, 25 transcripts) had the highest number of transcripts. Among the total species, the highest homology of *S. khuzistanica* and *S. rechingeri*'s TFs genes were with *S. miltiorrhiza*'s TFs genes (91 transcripts, 40%) (S4B Fig).

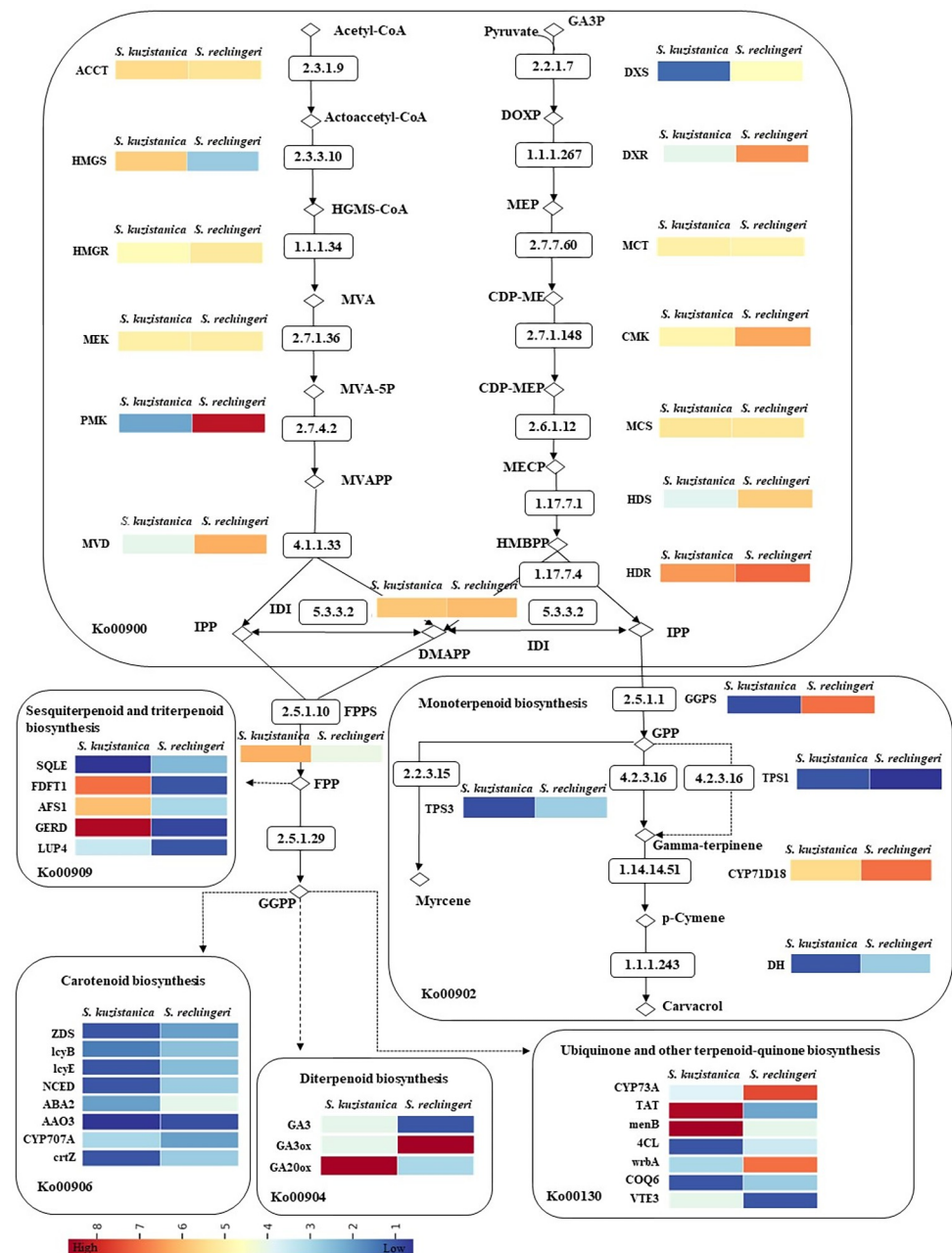


Fig 6. Expression patterns of *S. khuzistanica* and *S. rechingeri* transcripts involved in the terpenoid biosynthetic pathways. Logarithm of the FPKM value of all transcripts related to each gene was used. Ko numbers show the KEGG maps code related to each pathway. IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; GGPP, geranylgeranyl pyrophosphate; GGPS, GGPP synthase; GPP, geranyl pyrophosphate; FPP, farnesyl pyrophosphate; FPS, FPP synthase; TPS1, terpene synthase 1 (R-linalool synthase); TPS2, terpene synthase 2 ((4S)-limonene synthase); TPS3, terpene synthase 3; DH, carveol dehydrogenase; GA3, ent-kaurene oxidase; GA3ox, gibberellin 3beta-dioxygenase; GA20ox, gibberellin-44 dioxygenase; SQL, squalene monooxygenase; FDFT1, farnesyl-diphosphate farnesyltransferase; AFS1, alpha-farnesene synthase; GERD, (-)-germacrene D synthase; LUP4, beta-amyryn synthase; ZDS, zeta-carotene desaturase; lcyB, lycopene beta-cyclase; lcyE, lycopene epsilon-cyclase; NCED, 9-cis-epoxycarotenoid dioxygenase; ABA2, xanthoxin dehydrogenase; AAO3, abscisic-aldehyde oxidase; CYP707A, (+)-abscisic acid 8'-hydroxylase; crtZ, beta-carotene 3-hydroxylase; CYP73A, trans-cinnamate 4-monooxygenase; TAT, tyrosine aminotransferase; menB, naphthoate synthase; 4CL, 4-coumarate-CoA ligase; wrbA, NAD(P)H dehydrogenase; COQ6, ubiquinone biosynthesis monooxygenase; VTE3, MPBQ/MSBQ methyltransferase.

<https://doi.org/10.1371/journal.pone.0281351.g006>

Table 6. Differentially expressed gene family members which involved in terpenoid biosynthesis in *Satureja*.

Gene family	Differentially expressed genes
Cytochrome P450 (CYP450)	84
Terpene synthase (TPS)	11
Dehydrogenase (DH)	140
Transcription factor (TF)	226

<https://doi.org/10.1371/journal.pone.0281351.t006>

In order to understand the regulatory mechanism of terpenoid accumulation patterns, we analyzed the co-expression profiles between differentially expressed TFs and differentially expressed terpene biosynthesis-related enzyme genes (Fig 7) and their expression pattern as shown in S5 Fig.

As a result, ARF, MYC, bHLH and ERF family genes were predicted to be closely related to terpene biosynthesis genes (Fig 7, S12 Table). The results revealed that HDR was co-expressed with ARF2, ERF2 and bHLH13 genes and HDS was co-expressed with ARF2 and ARF11 genes. MVD was co-expressed with bHLH74. HMGR and TPS3 were co-expressed with MYC2 while CYP71D18 was co-expressed with bHLH13 gene. GGPS was only co-expressed with MYC4. These results highlighted a large number of candidate genes that regulate the accumulation of terpenoids in *Satureja*.

Analysis of essential oil components. Essential oils from leaves of *S. khuzistanica* and *S. rechingeri* were analyzed (S13 Table). A total of 24 compounds were identified in the essential oil of *S. khuzistanica* (sum of approximately 98.96%) and *S. rechingeri* (sum of approximately 98.18%). Phytochemical analyses of the leaf tissue showed differences in the amount of the main components of the essential oils. It can be due to the terpene synthase, other secondary metabolite, and their regulator genes (S14 Table).

Essential oil profiling by GC analysis represented that hydrocarbon monoterpenes with carvacrol had the highest percentage, followed by p-cymene γ -terpinene (Fig 8), and the percentage of other compounds (borneol, alpha-humulene, and bete-bisabolene) was very low. Also, the results of the variance analysis of the data showed that the difference between the two species of savory is significant at the level of 5% (Table 7).

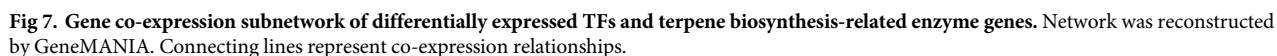
Verification of RNA-Seq data by qRT-PCR

In this study, since carvacrol is a major compound of *Satureja* essential oil, the genes involved in this pathway have received more attention. The expression patterns of six DEGs involved in the carvacrol biosynthesis in *Satureja* were validated by using qRT-PCR. As it is shown in Fig 9, the expression value of all six genes is highly consistent and in harmony with their expression from the RNA-Seq results ($R^2 = 0.99$). It is important to consider that the slight difference in obtained expression values from the qRT-PCR method with the RNA-Seq results is little due to different parameters such as the primers efficiency, quality of the PCR kit, and the integrity of the thermocycler.

Discussion

De novo transcriptome assembly

In this study, we report the first analysis of transcriptome data from the *Satureja* genus. The transcriptomes N50 length was higher than that of other members of this family such as *S. miltiorrhiza* (1340 bp) [61] and *D. tanguticum* (1118 bp) [27]. The annotation result indicated that among the total species, the high homology of *S. khuzistanica* and *S. rechingeri*'s assembled transcripts with *S. miltiorrhiza*; this could be due to the similarity of their family (Lamiaceae) [62].



The expression of genes involved in the terpenes biosynthesis pathway

PLOS ONE | <https://doi.org/10.1371/journal.pone.0281351> July 7, 2023

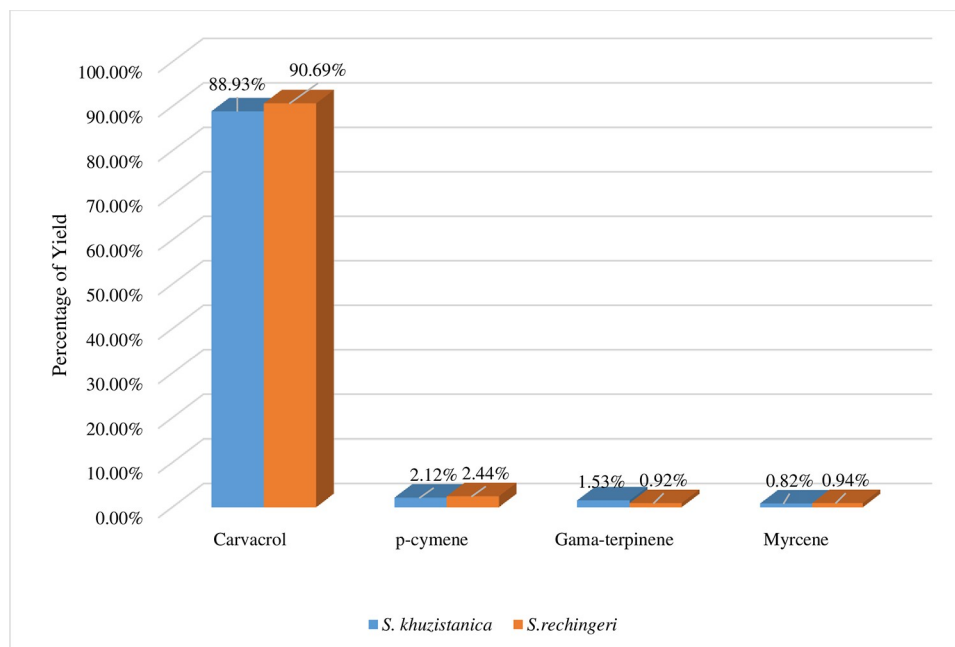


Fig 8. Yield percentage of essential oil's main components of *S. khuzistanica* and *S. rechingeri*.

<https://doi.org/10.1371/journal.pone.0281351.g008>

the MEP pathway of *S. rechingeri* than in *S. khuzistanica*. The present study suggested that the genes involved in the MEP pathway showed higher expression than the genes involved in the MVA pathway in both species. Both the MVA and MEP pathways were involved in carvacrol production, while the MEP pathway probably played a major role in the biosynthesis of carvacrol [63]. This proposed pathway has some similarities to that of carvacrol biosynthesis in oregano and thyme, in which the carvacrol was biosynthesized from IPP and DMADP which were derived via the MEP pathway [64].

The expression of genes involved in the monoterpene carvacrol biosynthesis pathway

The gene families of terpene synthases (TPS), cytochrome P450s (CYP450s), dehydrogenase (DHs), and transcription factors (TFs) may be involved in the biosynthesis of carvacrol [65]. The presence of TPS, CYP71D18, HDS, and GPPS genes in the DEG list, might be one of the causes of the increase in the carvacrol amount in *S. rechingeri* (S4 Table). In addition, an oregano *OvTPS2* gene encoding a protein with 91% amino acid identity to *Satureja* TPS2 appeared to be highly expressed in tissues producing high amounts of carvacrol [59]. Previous studies

Table 7. The pairwise comparison for the main essential oil components of *Satureja*.

Main component	(I) Species	(J) Species	Mean Difference (I-J)	Std. Error	Sig.b	95% confidence interval for difference	
						Lower Bound	Upper Bound
Carvacrol	<i>S. rechingeri</i>	<i>S. khuzistanica</i>	3.360*	0.885	0.019	0.903	5.817
Myrcene	<i>S. rechingeri</i>	<i>S. khuzistanica</i>	0.13	0.12	0.339	-0.202	0.462
p-cymene	<i>S. rechingeri</i>	<i>S. khuzistanica</i>	1.000*	0.233	0.013	0.353	1.647
Gama-terpinene	<i>S. rechingeri</i>	<i>S. khuzistanica</i>	0.605*	0.16	0.019	0.16	1.049

*Values are significant at P-value cut-off of < 0.05.

<https://doi.org/10.1371/journal.pone.0281351.t007>

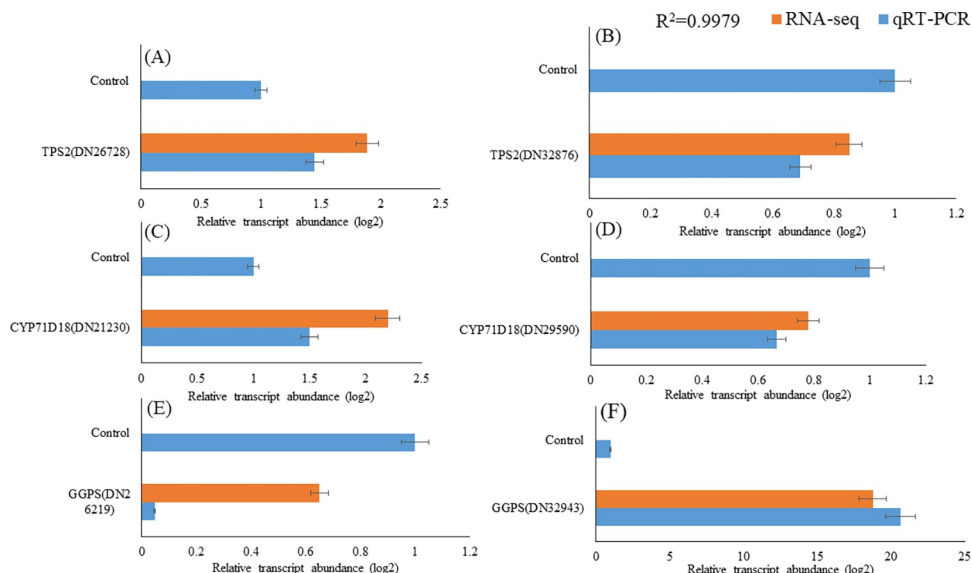


Fig 9. Validation of selected genes using qRT-PCR. (A) TPS2 (DN26728), terpene synthase 2 (B) TPS2 (DN32876), terpene synthase 2 (C) CYP71D18 (DN21230), cytochrome P450 71D18 (D) CYP71D18 (DN29590), cytochrome P450 71D18 (E) GGPS (DN26219), GGPP synthase (F) GGPS (DN32943), GGPP synthase.

<https://doi.org/10.1371/journal.pone.0281351.g009>

showed that the terpene synthase genes of *O. vulgare* appear to play a significant role in controlling terpene compounds. This suggests that terpene synthase expression levels directly control the compounds of the essential oil [60]. The TPS2 is mainly responsible for catalyzing the biosynthesis of monoterpenoids [66]. Similarly, *TvTPS2* transcript levels correlated positively with the high-carvacrol and -thymol levels in *Thymus* [67]. The terpene synthase genes can be utilized as markers for the breeding of medicinal plants with higher-value essential oil [68]. P450 family members are very important into the production of terpenoid secondary metabolites (monoterpenoids, diterpenoids, triterpenoids, and sesquiterpenoids) in plants [69]. The role of CYP450 family members in the biosynthesis of secondary metabolites compounds in medicinal plants such as *S. miltiorrhiza* [70], *Taxus chinensis* [71], *Catharanthus roseus* [72], *Medicago truncatula* [73] and *Panax ginseng* [74] has also been mentioned. Transcriptomic analysis of *Satureja* lead to the identification of CYP71D18, which exhibits 90.1% amino acid identity with *OvCYP71D181* and was well correlated with carvacrol levels. In the study of Crocoll et al. (2010) on the *O. vulgare* L., the role and involvement of CYP71D180 and CYP71D181 in the production of thymol and carvacrol were reported [60]. In *Mentha*, the role of P450 monooxygenases was confirmed in the biosynthesis of menthol and carvacrol [75].

Terpenoid biosynthetic regulation

One of the most important ways to control gene expression is transcriptional regulation that TFs play a critical role in it [76]. They regulate diverse biological processes such as growth, development, and secondary metabolism [77]. Several members of AP2/ERF, bHLH, MYB, WRKY [76, 78], NAC, DOF, bZIP, and HD-ZIP family proteins have been reported as regulators of secondary metabolite biosynthesis [79–81]. A positive correlation between the expression of differentially expressed TFs and the levels of differentially expressed terpene biosynthesis-related enzyme genes was also found. It has become evident that Both the MVA and MEP pathways are heavily regulated [82]. Basic bHLH TFs play a pivotal role in the biosynthesis of secondary metabolites [83]. MYC family members are bHLH TFs [84]. Some

MYC TFs control terpenoid biosynthesis in plants [78]. In lavender, *LaMYC4* was involved in the regulation of terpenoids and affects the expression of HMGR, DXS, DXR, FPPS, and GPPS in the terpenoid synthesis pathway. In addition, DXR expression was strongly associated with the expression of *LaMYC4* [85]. Previous transcriptome studies highlighted the close relationship between the silencing of NAC2 and the up-regulation of the expression levels of HMGR and DXS [86]. HMGR and DXS played important roles in the regulation of MVA and MEP pathways, respectively [87, 88].

Phytochemical analyses

Phytochemical analyses of the leaf tissue showed carvacrol content in *S. rechingeri* increased vs *S. khuzistanica*. Gama-terpinene, the second important compound of the savory plant, showed a significant difference between the two species. This is in agreement with the previous studies of essential oil components [56]. The regulation pattern of genes involved in terpenoids biosynthesis is probably due to the different genetic backgrounds of the two species. This data corresponded to the higher content of monoterpenoid in *S. rechingeri* and was confirmed by qRT-PCR.

Taken together, the results of previous studies confirm that the difference in the expression level of genes involved in terpenoid biosynthesis and identified transcription factors can be the cause of might be one of the causes of changing in the secondary metabolite biosynthesis amount in *S. khuzistanica* and *S. rechingeri*.

Conclusion

This study reported the results of comparative transcriptome analysis and key genes for investigation of the production of monoterpenoid in *Satureja*. A total of 60% of transcripts from all the libraries were annotated. Two species were evaluated for essential oil composition. Several transcripts encoding key enzymes involved in the MVA and MEP biosynthetic pathways were identified, which produce precursors for carvacrol biosynthesis. In total, 7026 DEGs were expressed in the different pattern in *S. rechingeri* compared to *S. khuzistanica*. A total of 461 DEGs involved in terpenoid biosynthesis were identified. The expression pattern of transcripts involved in the monoterpene biosynthetic pathway was evaluated. The expression pattern of 19 differentially expressed TFs genes was identified as pathway regulatory genes. These data will facilitate further study of the molecular mechanisms underlying monoterpenoid biosynthesis and gives rise to valuable resource data to explore carvacrol biosynthesis in the future.

Supporting information

S1 Fig. The number of transcripts with open reading frames in each species identified using the TransDecoder pipeline. Complete ORF sequences in which the first codon and the stop codon are present; 5 prime: sequences that contain the start codon but lack the stop codon; 3 prime: partial ORF sequences that lack the start codon; internal: sequences that lack both the start and stop codons.

(TIF)

S2 Fig. Top 10 hits of *S. khuzistanica* and *S. rechingeri*'s domain families.

(TIF)

S3 Fig. Phylogenetic analysis of Cytochrome P450 71D181 and terpene synthesis from *Satureja*. A) Cytochrome P450 71D181. The full name of the sequences are *SkCYP71D181*, Cytochrome P450 71D181 [*Saturejakhuzistanica*]; *SrCYP71D181*, Cytochrome P450 71D181 [*Satureja rechingeri*]; *TvL3H*, limonene-3-hydroxylase [*Thymus vulgaris*]; *TvL6H*, limonene-

6-hydroxylase [*Thymus vulgaris*]; MgCYP71D18, (-)-(4S)-Limonene-6-hydroxylase [*Mentha x gracilis*]; MgCYP71D94, Cytochrome P450 71D94 [*Mentha x gracilis*]; McL6H, limonene-6-hydroxylase [*Mentha canadensis*]; Mp(-)-L6H, (-)-limonene 6-hydroxylase-like cytochrome p450-dependent oxygenase [*Mentha x piperita*], MgCYP71D95, 71D95: (-)-(4S)-Limonene-3-hydroxylase [*Mentha x gracilis*]; MsCYP71D95, 71D95: Limonene-3-hydroxylase [*Mentha spicata*]; MpCYP71D15, (-)-(4S)-Limonene-3-hydroxylase; AltName: Full = Cytochrome P450 isoform PM2 [*Mentha x piperita*]; McL3H, (-)P450 limonene-3-hydroxylase [*Mentha canadensis*]; MaLH, limonene hydroxylase [*Mentha arvensis*]; Ms(-)-L3H, (-)-limonene 3-hydroxylase-like cytochrome p450-dependent oxygenase [*Mentha spicata*], MpCYP71D13, Cytochrome P450 71D13; AltName: Full = (-)-(4S)-Limonene-3-hydroxylase; SiCYP71D95-like, P450 71D95-like [*Sesamum indicum*]; SmCYP71D374; cytochrome P450 [*Salvia miltiorrhiza*] B) terpene synthesis. The full name of the sequences are TmTPS, gamma-terpinene synthase [*Thymus migricus*]; TtTPS gamma-terpinene synthase [*Thymus trautvetteri*]; TpTPS, gamma-terpinene synthase [*Thymus pubescens*]; TsTPS, putative gamma-terpinene synthase [*Thymus serpyllum*]; TdTPS, gamma-terpinene synthase [*Thymus daenensis*]; TvTPS2, terpene synthase 2 [*Thymus vulgaris*]; TcTPS, gamma-terpinene synthase [*Thymus caespitius*]; TfTPS, gamma-terpinene synthase [*Thymus fedtschenkoi*]; TkTPS, gamma-terpinene synthase [*Thymus kotschyanus*]; OvTPS2, terpene synthase 2 [*O. vulgare*]; SkTPS2, terpinene synthase [*Satureja khuzistanica*]; SrTPS2, terpinene synthase, [*Satureja rechingeri*]; OvTPS5, terpene synthase 5 [*O. vulgare*]; PvTPS10, terpene synthase 10 [*Prunella vulgaris*]; SrPS, pinene synthase [*Salvia rosmarinus*] The evolutionary history was inferred using the Neighbor-Joining method. Evolutionary analyses were conducted in MEGA7. The bootstrap replications were set to 100.

(TIF)

S4 Fig. Transcription factor genes of *Satureja*. A) Percent of identified TFs families B) Percent of *Satureja* TFs genes with high homology of plant species.

(TIF)

S5 Fig. Heatmap of differentially expressed TFs families. Average of FPKM of all transcripts related to each TF gene was used. Ko numbers show the KEGG maps code related to monoterpene biosynthetic pathway. Specie ID of each replicate is Sk_1: *S. khuzistanica*_1, Sk_2: *S. khuzistanica*_2, Sk_3: *S. khuzistanica*_3, Sr_1: *S. rechingeri*_1, Sr_2: *S. rechingeri*_2, Sr_3: *S. rechingeri*_3.

(TIF)

S1 Table. Specific oligonucleotides for qRT-PCR analysis.

(XLSX)

S2 Table. Trinity assembly stats report of assembled contigs.

(XLSX)

S3 Table. Summary of functional annotation result.

(XLSX)

S4 Table. Differentially expressed and annotation of total transcripts in *S. kuzistanica* and *S. rechingeri*.

(XLSX)

S5 Table. GO enrichment analysis for DEGs transcripts in *Satureja*.

(XLSX)

S6 Table. GO enrichment analysis for DEGs transcripts related to the terpenoid biosynthesis in *Satureja*.

(XLSX)

S7 Table. Number of DEGs were assigned to KEGG pathways.

(XLSX)

S8 Table. DEGs encoding enzymes involved in terpenoid biosynthesis in *Satureja*.

(XLSX)

S9 Table. Transcripts related to biosynthesis of other secondary metabolites in *S. khuzistanica* and *S. rechingeri* leaf transcriptome defined by KAAS.

(XLSX)

S10 Table. Transcripts related to terpenoid backbone biosynthesis in *S. khuzistanica* and *S. rechingeri* leaf transcriptome defined by KAAS.

(XLSX)

S11 Table. Differentially expressed and annotation for transcription factor gene families of *Satureja*.

(XLSX)

S12 Table. Co-expression network among differentially expressed TFs and terpene biosynthesis-related enzyme genes.

(XLSX)

S13 Table. GC profile of *S. khuzistanica* and *S. rechingeri* leaf essential oils.

(XLSX)

S14 Table. The two-way ANOVA for the essential oil components experiment of the leaf tissue of *Satureja* (complete model).

(XLSX)

Author Contributions

Conceptualization: Ahmad Ismaili, Farhad Nazarian Firouzabadi, Hasan Mumivand.

Data curation: Somayeh Shams, Ahmad Ismaili.

Formal analysis: Somayeh Shams.

Funding acquisition: Ahmad Ismaili.

Investigation: Somayeh Shams.

Methodology: Ahmad Ismaili.

Project administration: Ahmad Ismaili.

Resources: Somayeh Shams, Hasan Mumivand.

Software: Somayeh Shams, Karim Sorkheh.

Supervision: Ahmad Ismaili.

Validation: Somayeh Shams, Ahmad Ismaili, Hasan Mumivand.

Visualization: Ahmad Ismaili, Farhad Nazarian Firouzabadi, Hasan Mumivand.

Writing – original draft: Somayeh Shams.

Writing – review & editing: Ahmad Ismaili, Farhad Nazarian Firouzabadi.

References

1. Yang W, Chen X, Li Y, Guo S, Wang Z, Yu X. Advances in pharmacological activities of terpenoids. *Nat Prod Commun.* 2020; 15(3):1934578X20903555. <https://doi.org/10.1177/1934578X20903555>.
2. Nagegowda DA. Plant volatile terpenoid metabolism: biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Lett.* 2010; 584(14): 2965–2973. <https://doi.org/10.1016/j.febslet.2010.05.045> PMID: 20553718
3. Wang Q, Quan S, Xiao H. Towards efficient terpenoid biosynthesis: manipulating IPP and DMAPP supply. *Bioresour Bioprocess.* 2019; 6(1): 1–3. <https://doi.org/10.1186/s40643-019-0242-z>.
4. Croteau R, Kutchan TM, Lewis NG. Natural products (secondary metabolites). *Physiol Mol Biol Plants.* 2000; 24:1250–319.
5. Zielińska-Blajet M, Feder-Kubis J. Monoterpenes and their derivatives—Recent development in biological and medical applications. *Int. J. Mol. Sci.* 2020;25; 21(19):7078. <https://doi.org/10.3390/ijms21197078> PMID: 32992914
6. Jamzad Z. *Thymus* and *Satureja* species of Iran. Publications of research institute of forests and rangelands; 2009. pp. 171.
7. Raina AP, Negi KS. Essential oil composition of *Origanum majorana* and *O. vulgare* ssp. *hirtum* growing in India. *Chem. Nat. Compd.* 2012; 47(6): 1015–1017. <https://doi.org/10.1007/s10600-012-0133-4>.
8. Chograni H, Zaouali Y, Rajeb C, Boussaid M. Essential oil variation among natural populations of *Lavandula multifida* L. (Lamiaceae). *Chem Biodivers.* 2010; 7(4): 933–42. <https://doi.org/10.1002/cbdv.200900201> PMID: 20397226
9. Ulukanli Z, Cenet M, Ince H, Yilmaztekin M. Antimicrobial and herbicidal activities of the essential oil from the Mediterranean *Thymus eigii*. *J Essent. Oil-Bear Plants.* 2018; 21(1): 214–222. <https://doi.org/10.1080/0972060X.2017.1327824>.
10. Zavatti M, Zanoli P, Benelli A, Rivasi M, Baraldi C, Baraldi M. Experimental study on *Satureja montana* as a treatment for premature ejaculation. *J Ethnopharmacol.* 2011; 133(2): 629–33. <https://doi.org/10.1016/j.jep.2010.10.058>.
11. Jamzad Z. A new species of the genus *Satureja* (Labiatae) from Iran. *Iran J Bot.* 1994; 6: 215–8.
12. Sefidkon F, Emami Bistgani Z. Integrative review on ethnobotany, essential oil, phytochemical, agronomy, molecular and pharmacological properties of *Satureja* species. *J. Essent Oil Res.* 2021; 33(2): 114–32. <https://doi.org/10.1080/10412905.2021.1885512>.
13. Hadian J, Akramian M, Heydari H, Mumivand H, Asghari B. Composition and in vitro antibacterial activity of essential oils from four *Satureja* species growing in Iran. *Nat Prod Res.* 2012; 26(2): 98–108. <https://doi.org/10.1080/14786419.2010.534734>.
14. Babajafari S, Nikaein F, Mazloomi SM, Zibaeenejad MJ, Zargarani A. A review of the benefits of *Satureja* species on metabolic syndrome and their possible mechanisms of action. *Journal of evidence-based complementary & alternative medicine.* 2015; 20(3):212–23. <https://doi.org/10.1177/2156587214564188> PMID: 25563729
15. Zarrin M, Amirrajab N, Sadeghi-Nejad B. In vitro antifungal activity of *Satureja khuzestanica* Jamzad against *Cryptococcus neoformans*. *Pak J Med Sci.* 2010; 26(4): 880–882.
16. Hadian J, Hossein Mirjalili M, Reza Kanani M, Salehnia A, Ganjipoor P. Phytochemical and morphological characterization of *Satureja khuzestanica* Jamzad populations from Iran. *Chem Biodivers.* 2011; 8(5): 902–915. <https://doi.org/10.1002/cbdv.201000249> PMID: 21560239
17. Khaledi A, Meskini M. A systematic review of the effects of *Satureja khuzestanica* Jamzad and *Zataria multiflora* Boiss against *Pseudomonas aeruginosa*. *Iran J Med Sci.* 2020; 45: 83. <https://doi.org/10.30476/IJMS.2019.72570>. PMID: 32210484
18. Ghodrati L, Alizadeh A, Ketabchi S. Essential oil constituents and antimicrobial activities of Iranian *Satureja khuzestanica* Jamzad. *Int J Biosci.* 2015; 6(3): 249257. <http://dx.doi.org/10.12692/ijb/6.3.249-257>.
19. Khojasteh A, Mirjalili MH, Alcalde MA, Cusido RM, Eibl R, Palazon J. Powerful plant antioxidants: a new biosustainable approach to the production of rosmarinic acid. *Antioxidants.* 2020; 9(12): 1273. <https://doi.org/10.3390/antiox9121273> PMID: 33327619
20. Dragalin I, Aricu A, Ciocarlan N, Ciocarlan A, Codita V. GC-MS analysis of the essential oil of *Satureja subspicata* Bartl. ex Vis. of Moldovan origin. *Chem J Mold.* 2016; 11(2): 105–108. [https://dx.doi.org/10.19261/cjm.2016.11\(2\).06](https://dx.doi.org/10.19261/cjm.2016.11(2).06).

21. Shanaida M, Ivanusa I, ernychna I. Phytochemical analysis of secondary metabolites of *Satureja hortensis* L. Int J Pharm Pharm Sci. 2017; 9(2): 315–318. <https://doi.org/10.22159/ijpps.2017v9i2.15923>.
22. Anamika K, Verma S, Jere A, Desai A. Transcriptomic profiling using next generation sequencing—advances, advantages, and challenges. Next generation sequencing—advances, applications and challenges; 2016; 9: pp. 7355–65. <https://doi.org/10.5772/61789>.
23. Feng S, Zhao L, Liu Z, Liu Y, Yang T, Wei A. De novo transcriptome assembly of *Zanthoxylum bungeanum* using Illumina sequencing for evolutionary analysis and simple sequence repeat marker development. Sci Rep. 2017; 7(1): 1–11. <https://doi.org/10.1038/s41598-017-15911-7>.
24. Delli-Ponti R, Shivhare D, Mutwil M. Using gene expression to study specialized metabolism—a practical guide. Front Plant Sci. 2021; 11: 2074. <https://doi.org/10.3389/fpls.2020.625035> PMID: 33510763
25. Ma XH, Ma Y, Tang JF, He YL, Liu YC, Ma XJ, et al. The biosynthetic pathways of tanshinones and phenolic acids in *Salvia miltiorrhiza*. Molecules. 2015; 20(9): 16235–16254. <https://doi.org/10.3390/molecules200916235> PMID: 26370949
26. Yang L, Ding G, Lin H, Cheng H, Kong Y, Wei Y, et al. 2013. Transcriptome analysis of medicinal plant *S. miltiorrhiza* and identification of genes related to tanshinone biosynthesis. PLoS One. 2013; 8(11): e80464. <https://doi.org/10.1371/journal.pone.0080464>.
27. Li H, Fu Y, Sun H, Zhang Y, Lan X. Transcriptomic analyses reveal biosynthetic genes related to rosmarinic acid in *Dracocephalum tanguticum*. Sci Rep. 2017; 7: 74. <https://doi.org/10.1038/s41598-017-00078-y> PMID: 28250432
28. Meena S, Kumar SR, Venkata Rao D, Dwivedi V, Shilpashree H, Rastogi S, et al. De novo sequencing and analysis of lemongrass transcriptome provide first insights into the essential oil biosynthesis of aromatic grasses. Front Plant Sci. 2016; 7: 1129. <https://doi.org/10.3389/fpls.2016.01129>.
29. Zhang SD, Ling LZ. De novo sequencing and comparative analysis of peduncle transcriptomes in *Hovea acerba*. Agri Gene. 2017; 3: 32–36. <https://doi.org/10.1016/j.aggene.2016.11.001>
30. Liu Y, Wang Y, Guo F, Zhan L, Mohr T, Cheng P, et al. Deep sequencing and transcriptome analyses to identify genes involved in secoiridoid biosynthesis in the Tibetan medicinal plant *Swertia mussotii*. Sci Rep. 2017; 7(1): 43108. <https://doi.org/10.1038/srep43108> PMID: 28225035
31. Hao C, Xia Z, Fan R, Tan L, Hu L, Wu B, et al. De novo transcriptome sequencing of black pepper (*Piper nigrum* L.) and an analysis of genes involved in phenylpropanoid metabolism in response to *Phytophthora capsici*. BMC Genet. 2016; 17(1): 1–4. <https://dx.doi.org/10.1186/s12864-016-3155-7>.
32. Alizadeh A. Essential oil composition, phenolic content, antioxidant, and antimicrobial activity of cultivated *Satureja rechingeri* Jamzad at different phenological stages. Naturforsch., C, J. Biosci. 2015; 70(3–4), 51–58. <https://doi.org/10.1515/znc-2014-4121>.
33. Ahmadi SH, Sefidkon F, Babakhanlo P, Asgari FA, Khademi K, Karimifar MA. Comparing essential oil composition of *Satureja bachtiarica* Bunge before and full flowering stages in field and provenance. IJMAPR. 2009; 25(2):159–69. <https://dx.doi.org/10.22092/ijmapr.2009.7187>.
34. KiZil S, Ipek A, Arslan N, Khawar KM. Effect of different developing stages on some agronomical characteristics and essential oil composition of oregano (*Origanum onites*). J. Crop Hortic. Sci. 2008; 36(1):71–6. <https://doi.org/10.1080/01140670809510222>.
35. Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom. 2010. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
36. Bolger A. M, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 2014; 30(15): 2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
37. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013; 8(8): 1494–1512. <https://doi.org/10.1038/nprot.2013.084> PMID: 23845962
38. Schliesky S, Gowik U, Weber AP, Bräutigam A. RNA-seq assembly—are we there yet?. Front Plant Sci. 2012; 3: 220. <https://doi.org/10.3389/fpls.2012.00220> PMID: 23056003
39. Consortium U. Activities at the universal protein resource (UniProt). Nucleic Acids Res. 2014; 42(D1): D191–198. <https://doi.org/10.1093/nar/gkt1140> PMID: 24253303
40. Muller J, Szklarczyk D, Julien P, Letunic I, Roth A, Kuhn M, et al. eggNOG v2. 0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. Nucleic Acids Res. 2009; 38(suppl_1): D190–D195. <https://doi.org/10.1093/nar/gkp951>.
41. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007; 35(suppl_2): W182–W185. <https://doi.org/10.1093/nar/gkm321> PMID: 17526522

42. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein family's database: towards a more sustainable future. *Nucleic Acids Res*. 2016; 44(D1): D279–285. <https://doi.org/10.1093/nar/gkv1344>.
43. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform*. 2011; 12: 323. <https://doi.org/10.1186/1471-2105-12-323>.
44. Jiménez-Jacinto V, Sanchez-Flores A, Vega-Alvarado L. Integrative differential expression analysis for multiple experiments (IDEAMEX): a web server tool for integrated rna-seq data analysis. *Front Genet*. 2019; 10: 279. <https://doi.org/10.3389/fgene.2019.00279> PMID: 30984248
45. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007. <http://bioinfoqpb.cnb.csic.es/tools/venny/index.html>.
46. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*. 2006; 34(suppl_2): W293–W297. <https://doi.org/10.1093/nar/gkl031>.
47. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019; 47 (W1): W191–W198. <https://doi.org/10.1093/nar/gkz369> PMID: 31066453
48. Deng W, Wang Y, Liu Z, Cheng H, Xue Y, Heml: a toolkit for illustrating heat maps. *PloS One*. 2014; 9 (11):e111988. <https://doi.org/10.1371/journal.pone.0111988> PMID: 25372567
49. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE*. 2011. <https://doi.org/10.1371/journal.pone.0021800> PMID: 21789182
50. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res*. 2016; gkw982. <https://doi.org/10.1093/nar/gkw982>.
51. Fernandez NF, Gundersen GW, Rahman A, Grimes ML, Rikova K, Hornbeck P, et al. Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data. *Sci. Data*. 2017; 10; 4(1):1–2. <https://doi.org/10.1038/sdata.2017.151> PMID: 28994825
52. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010; 1; 38(suppl_2): W214–220. <https://doi.org/10.1093/nar/gkq537> PMID: 20576703
53. Xiao X, Ma J, Wang J, Wu X, Li P, Yao Y. Validation of suitable reference genes for gene expression analysis in the halophyte *Salicornia europaea* by real-time quantitative PCR. *Front Plant Sci*. 2015; 5: 788. <https://doi.org/10.3389/fpls.2014.00788>.
54. Pfaffl MW. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res*. 2001; 29: e45–e45. <https://doi.org/10.1093/nar/29.9.e45> PMID: 11328886
55. Pfaffl MW, Horgan GW, Dempfle L. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res*. 2002; 30 (9): e36–e36. <https://doi.org/10.1093/nar/30.9.e36> PMID: 11972351
56. Nooshkam A, Mumivand H, Hadian J, Alemardan A, Morshedloo MR. Drug yield and essential oil and carvacrol contents of two species of *Satureja* (*S. khuzistanica* Jamzad and *S. rechingeri* Jamzad) cultivated in two different locations. *J Appl Res Med Aromat Plants*. 2017; 6: 126–130. <https://doi.org/10.1016/j.jarmap.2017.04.002>.
57. Morshedloo MR, Mumivand H, Craker LE, Maggi F. Chemical composition and antioxidant activity of essential oils in *O. vulgare* subsp. *gracile* at different phenological stages and plant parts. *J Food Process Preserv*. 2018; 42(2): e13516. <https://doi.org/10.1111/jfpp.13516>.
58. Gierse RM, Reddem ER, Alhayek A, Baitinger D, Hamid Z, Jakobi H, et al. Identification of a 1-deoxy-D-xylulose-5-phosphate synthase (DXS) mutant with improved crystallographic properties. *Biochem Biophys Res Commun*. 2021; 539: 42–47. <https://doi.org/10.1016/j.bbrc.2020.12.069> PMID: 33421767
59. Crocoll C, Asbach J, Novak J, Gershenzon J, Degenhardt J. Terpene synthases of oregano (*O. vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. *Plant Mol Biol*. 2010; 73(6): 587–603. <https://doi.org/10.1007/s11103-010-9636-1>.
60. Crocoll C. Biosynthesis of the phenolic monoterpenes, thymol and carvacrol, by terpene synthases and cytochrome P450s in oregano and thyme (Doctoral dissertation, Thüringer Universitäts-und Landesbibliothek Jena). 2010:1–51.
61. Zhou W, Huang Q, Wu X, Zhou Z, Ding M, Shi M, et al. Comprehensive transcriptome profiling of *S. mil-tiorrhiza* for discovery of genes associated with the biosynthesis of tanshinones and phenolic acids. *Sci Rep*. 2017; 7(1): 1–2. <https://doi.org/10.1038/s41598-017-10215-2>.
62. Uritu CM, Mihai CT, Stanciu GD, Dodi G, Alexa-Stratulat T, Luca A, et al. Medicinal plants of the family Lamiaceae in pain therapy: A review. *Pain Res Manag*. 2018. <https://doi.org/10.1155/2018/7801543> PMID: 29854039

63. Ramak P, Osaloo SK, Ebrahimzadeh H, Sharifi M, Behmanesh M. Inhibition of the mevalonate pathway enhances carvacrol biosynthesis and DXR gene expression in shoot cultures of *Satureja khuzistanica* Jamzad. J. Plant Physiol. 2013; 170(13):1187–93. <https://doi.org/10.1016/j.jplph.2013.03.013>.
64. Dudareva N, Andersson S, Orlova I, Gatto N, Reichelt M, Rhodes D, et al. The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. Proc. Natl. Acad. Sci. 2005; 102(3):933–8. <https://doi.org/10.1073/pnas.0407360102> PMID: 15630092
65. Soltani Howyze M, Sadat Noori SA, Shariati J V, Amiripour M. Comparative transcriptome analysis to identify putative genes involved in thymol biosynthesis pathway in medicinal plant *Trachyspermum ammi* L. Sci. Rep. 2018; 8(1):1–9. <https://doi.org/10.1038/s41598-018-31618-9>.
66. Chen Z, Vining KJ, Qi X, Yu X, Zheng Y, Liu Z, et al. Genome-wide analysis of terpene synthase gene family in *Mentha longifolia* and catalytic activity analysis of a single terpene synthase. Genes. 2021; 12(04):518. <https://doi.org/10.3390/genes12040518>.
67. Krause ST, Liao P, Crocoll C, Boachon B, Förster C, Leidecker F, et al. The biosynthesis of thymol, carvacrol, and thymohydroquinone in Lamiaceae proceeds via cytochrome P450s and a short-chain dehydrogenase. Proc. Natl. Acad. Sci. 2021; 118(52):e2110092118. <https://doi.org/10.1073/pnas.2110092118> PMID: 34930840
68. Novak J, Lukas B, Bolzer K, Grausgruber-Groger S, Degenhardt J. Identification and characterization of simple sequence repeat markers from a glandular *O. vulgare* expressed sequence tag. Mol Ecol Resour. 2008; 8(3): 599–601. <https://doi.org/10.1111/j.1471-8286.2007.02059.x>.
69. Mani V, Park S, Kim JA, Lee SI, Lee K. Metabolic Perturbation and Synthetic Biology Strategies for Plant Terpenoid Production—An Updated Overview. Plants. 2021; 10(10): 2179. <https://doi.org/10.3390/plants10102179> PMID: 34685985
70. Guo J, Zhou YJ, Hillwig ML, Shen Y, Yang L, Wang Y, et al. CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. Proc Natl Acad Sci. 2013; 110(29): 12108–12113. <https://doi.org/10.1073/pnas.1218061110> PMID: 23812755
71. Liao W, Zhao S, Zhang M, Dong K, Chen Y, Fu C, et al. Transcriptome assembly and systematic identification of novel cytochrome P450s in *Taxus chinensis*. Front Plant Sci. 2017; 8: 1468. <https://doi.org/10.3389/fpls.2017.01468>.
72. Besseau S, Kellner F, Lanoue A, Thamm A. M, Salim V, Schneider B, et al. A pair of tabersonine 16-hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in *Catharanthus roseus*. Plant Physiol. 2013; 163(4): 1792–803. <http://dx.doi.org/10.1104/pp.113.222828>.
73. Carelli M, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, et al. *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. Plant Cell. 2011; 23(8): 3070–3081. <https://doi.org/10.1105/tpc.111.087312>.
74. Han JY, Kim MJ, Ban YW, Hwang HS, Choi YE. The involvement of β -amyrin 28-oxidase (CYP716A52v2) in oleanane-type ginsenoside biosynthesis in *Panax ginseng*. Plant Cell Physiol. 2013; 54(12): 2034–2046. <https://doi.org/10.1093/pcp/pct141>.
75. Krause S. Biosynthesis of oxygenated monoterpenes in Thyme, Melaleuca, and Eucalyptus species. (Doctoral dissertation, Martin-Luther Universitäts). 2016.
76. Guo AY, Chen X, Gao G, Zhang H, Zhu QH, Liu XC, et al. PlantTFDB: a comprehensive plant transcription factor database. Nucleic Acids Res. 2007; 36(suppl_1): D966–9. <https://doi.org/10.1093/nar/gkm841> PMID: 17933783
77. Patra B, Schluttenhofer C, Wu Y, Pattanaik S, Yuan L. Transcriptional regulation of secondary metabolite biosynthesis in plants. Biochim Biophys Acta Gene Regul Mech. 2013; 1829(11): 1236–1247. <https://doi.org/10.1016/j.bbagrm.2013.09.006> PMID: 24113224
78. Yang CQ, Fang X, Wu XM, Mao YB, Wang LJ, Chen XY. Transcriptional Regulation of Plant Secondary Metabolism F J Integr Plant Biol. 2012; 54(10): 703–712. <https://doi.org/10.1111/j.1744-7909.2012.01161.x>.
79. Shen W, Li H, Teng R, Wang Y, Wang W, Zhuang J. Genomic and transcriptomic analyses of HD-Zip family transcription factors and their responses to abiotic stress in tea plant (*Camellia sinensis*). Genomics. 2019; 111(5): 1142–1151. <https://doi.org/10.1016/j.ygeno.2018.07.009> PMID: 30031053
80. Wu Q, Bai X, Zhao W, Shi X, Xiang D, Wan Y, et al. Investigation into the underlying regulatory mechanisms shaping inflorescence architecture in *Chenopodium quinoa*. BMC Genet. 2019; 20(1): 1–20. <https://doi.org/10.1186/s12864-019-6027-0>.
81. Li CF, Zhu Y, Yu Y, Zhao QY, Wang SJ, Wang XC, et al. Global transcriptome and gene regulation network for secondary metabolite biosynthesis of tea plant (*Camellia sinensis*). BMC Genom. 2015; 16(1):560.
82. Schrader J, Bohlmann J. Biotechnology of isoprenoids. Cham, Switzerland: Springer; 2015.

83. Mertens J, Pollier J, Vanden BR, López Vidriero I, Franco-Zorrilla JM, Goossens A. The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiol.* 2016; 170(1):194–210. <https://doi.org/10.1104/pp.15.01645>.
84. Huang D, Dai W. Molecular characterization of the basic helix–loop–helix (bHLH) genes that are differentially expressed and induced by iron deficiency in *Populus*. *Plant Cell Rep.* 2015; 34(7):1211–24. <https://doi.org/10.1007/s00299-015-1779-8> PMID: 25721202
85. Dong Y, Zhang W, Li J, Wang D, Bai H, Li H, et al. The transcription factor *LaMYC4* from lavender regulates volatile Terpenoid biosynthesis. *BMC Plant Biol.* 2022; 22(1):1–5. <https://doi.org/10.1186/s12870-022-03660-3>.
86. Zhang H, Xu J, Chen H, Jin W, Liang Z. Characterization of NAC family genes in *Salvia miltiorrhiza* and NAC2 potentially involved in the biosynthesis of tanshinones. *Phytochem.* 2021; 191:112932. <https://doi.org/10.1016/j.phytochem.2021.112932> PMID: 34454170
87. Cordoba E, Salmi M, León P. Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.* 2009; 60(10):2933–43. <https://doi.org/10.1093/jxb/erp190> PMID: 19584121
88. Kliebenstein DJ. Making new molecules-evolution of structures for novel metabolites in plants. *Curr Opin Plant Biol.* 2013; 16:112–117. <https://doi.org/10.1016/j.pbi.2012.12.004> PMID: 23295108