

Article

***De novo* Sequencing and Analysis of *Salvia hispanica* Tissue-Specific Transcriptome and Identification of Genes Involved in Terpenoid Biosynthesis**

James Wimberley ^{1,2}, Joseph Cahill ³ and Hagop S. Atamian ^{2,4,*}

¹ Computational and Data Sciences Program, Chapman University, Orange, CA 92866, USA; wimbe106@mail.chapman.edu

² Schmid College of Science and Technology, Chapman University, Orange, CA 92866, USA

³ Ventura Botanical Gardens, Ventura, CA 93001, USA; jcahill@venturabotanicalgardens.com

⁴ Biological Sciences Program, Chapman University, Orange, CA 92866, USA

* Correspondence: atamian@chapman.edu; Tel.: +1-(714)-289-2023

Received: 25 February 2020; Accepted: 19 March 2020; Published: 24 March 2020



Abstract: *Salvia hispanica* (commonly known as chia) is gaining popularity worldwide as a healthy food supplement due to its low saturated fatty acid and high polyunsaturated fatty acid content, in addition to being rich in protein, fiber, and antioxidants. Chia leaves contain plethora of secondary metabolites with medicinal properties. In this study, we sequenced chia leaf and root transcriptomes using the Illumina platform. The short reads were assembled into contigs using the Trinity software and annotated against the Uniprot database. The reads were de novo assembled into 103,367 contigs, which represented 92.8% transcriptome completeness and a diverse set of Gene Ontology terms. Differential expression analysis identified 6151 and 8116 contigs significantly upregulated in the leaf and root tissues, respectively. In addition, we identified 30 contigs belonging to the Terpene synthase (TPS) family and demonstrated their evolutionary relationships to tomato TPS family members. Finally, we characterized the expression of *S. hispanica* TPS members in leaves subjected to abiotic stresses and hormone treatments. Abscisic acid had the most pronounced effect on the expression of the TPS genes tested in this study. Our work provides valuable community resources for future studies aimed at improving and utilizing the beneficial constituents of this emerging healthy food source.

Keywords: RNA-seq; assembly; plant; differential expression

1. Introduction

Salvia hispanica L. (commonly known as chia) is an annual self-pollinated species that belongs to the mint family (Lamiaceae) and is native to central and southern Mexico and Guatemala [1]. *S. hispanica* grows up to six feet long and develops lush green foliage rich in essential oils before producing long purple or white flowers. These flowers develop to produce thousands of small (2 mm in length) highly nutritious edible seeds. *S. hispanica* has a long history of plant–human interactions. In pre-Columbian Mesoamerica, the plant was a major commodity, similar to bean, corn, and squash, and Aztecs valued its seeds for food, medicine, and oil [2]. The codices of 16th century Mexico provide a wealth of ethnobotanical information and indicate that large areas of agricultural land were devoted exclusively to chia cultivation [2]. However, after Spanish contact and colonization, the cultivation was prohibited due to its connection to Aztec cultural and religious rituals. Consequently, the plant was largely overlooked as a food crop until its re-emergence as an alternative crop and a health food in the beginning of the 20th century [1].

Chia seed provides a remarkably balanced and close to complete nutritional source with 34.4% total dietary fiber, 31% total lipids, 16% protein, 5.8% moisture, and high amounts (335–860 mg/100 g)

of calcium, phosphorus, potassium, and magnesium [1,3,4]. The oil content of chia seed (31%) is higher than that of other oilseeds of commercial importance, such as soybean (24%) and cotton-seed (24%) [4]. The fatty acids of chia seed oil are highly unsaturated, with their main components being linolenic (50–57%) and linoleic (17–26%) fatty acids. This represents the highest known percentage of linolenic fatty acid of any plant source [5]. Compared to the seed, chia leaf has 60% more palmitic acid content, but only 25% the concentration of α -Linolenic acid [6].

Although chia is better known for its seed's healthy nutritional composition, its leaves also provide a diverse source of metabolic products. According to the accounts of ethnobotanical use during the post-1600 AD period, vegetative plant parts were associated with medicinal uses [2]. Besides fatty acids, chia leaves contain essential oils that have the potential for commercial uses in the food flavoring and fragrance industry. These leaf essential oils also have antimicrobial properties [7] and could be used as biopesticides to protect plants from pathogen and insect attacks [8]. Gas chromatography–mass spectrometry analysis of the leaf oil composition from plants grown in southern California, southeastern Texas, and northwestern Argentina identified large number of components, of which the most abundant were sesquiterpenes β -caryophyllene, globulol, γ -muurolene, α -humulene, germacrene-B, and widdrol and the monoterpene β -pinene [9]. Similarly, an independent analysis of chia leaf oil constituents identified 60 different sesquiterpenes, accounting for 84.5% of the oil [7].

Terpenoids represent the most diverse group of plant secondary metabolites, with at least 25,000 compounds [10]. They are normally produced in vegetative tissues and flowers and are responsible for the distinct smells in plants [11]. Terpenoids are derived from the isomeric 5-carbon building blocks isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) through the methylerythritol 4-phosphate (MEP) and mevalonic acid (MVA) pathways [12]. The terpenoids released from plants as volatiles play important roles in plant-biotic interactions including defense against herbivores [13] and attraction of pollinators [14]. The vast diversity of the plant terpenoids is due to the action of enzymes encoded by the terpene synthase (TPS) gene family. *TPS* genes have been identified and characterized from a number of plant species and the size of the *TPS* family in the currently sequenced plant genomes ranges from 20 to 150 genes [12]. The *TPS* enzymes synthesize the backbone of the specialized monoterpenes, sesquiterpenes and diterpenes [15,16]. The plant *TPS* family members are divided into seven clades *TPSa-h* [12]. Within the angiosperm specific clades, the *TPS-a* clade mostly includes sesquiterpene and diterpene synthases while the *TPS-b* and *TPS-g* clades mostly include monoterpene synthases [12]. *TPS-c* clade is believed to be the ancestral clade and contains the gymnosperm and angiosperm *CPS* genes. *TPS-e/f* clade contains gymnosperm and angiosperm *KS* genes (in angiosperm and gymnosperm plants, ent-kaurene is also synthesized from GGPP via CPP in two steps, but the reactions are catalyzed by separate *CPS* and *KS* enzymes) and various other *TPSs* [16].

Arabidopsis and tomato genomes encode 32 and 29 potentially functional *TPS* genes, respectively. On the other hand, the *TPS* family has expanded in Grapevine (*Vitis vinifera* L.) coding for 69 putatively functional proteins [16–19]. Within the *TPS* family, sesquiterpene synthases catalyze the conversion of farnesyl diphosphate (FDP), generating a diverse array of sesquiterpene compounds [20]. Sesquiterpenes are hydrophobic bioactive compounds produced by plants that play important roles in defense against insects and pathogenic microorganisms [21]. Plant-produced sesquiterpenes are used as ingredients in pharmaceutical, cosmetic and flavoring products [22]. The chia leaf oil sesquiterpenes are mostly represented by sesquiterpene hydrocarbons (53.9%) and oxygenated sesquiterpenes (30.6%). Some abundant sesquiterpene hydrocarbons include (*Z*)-caryophyllene (11.5%), (*E*)-caryophyllene (10.6%), α -humulene (4.8%), δ -amorphenone (3.1%), and γ -gurjunene (3.1%). Oxygenated sesquiterpenes are more uniformly distributed with α -eudesmol (3.8%), caryophyllene oxide (2.7%), and spathulenol (2.2%) as the main representatives [7]. Monoterpenes constitute 0.4% of the chia leaf essential oil. The metabolic profile of chia leaves also includes several flavonoids and hydroxycinnamic acids such as apigenin and luteolin glycosides, aglycones quercetin methyl ether and naringenin, and quercetin- and kaempferol-based flavonoids [23].

RNA sequencing (RNA-Seq) is a powerful tool that is widely used in profiling the gene constituent of non-model species. The *de novo* sequencing and assembly of a transcriptome is the first step in gaining insights into the genes and molecular pathways underlying the different phenotypes in non-model plant species. In this study, we sequenced and assembled the *S. hispanica* leaf and root transcriptomes into 103,367 contigs with an estimated 92.2% completeness. Functional and Gene Ontology (GO) analysis of the assembled transcriptome identified diverse gene categories. Differential gene expression analysis identified 6151 and 8116 contigs that had higher expression in *S. hispanica* leaf and root, respectively. Genes encoding key enzymes involved in vitamin biosynthesis and homologs of terpene synthases were identified and their expression further characterized. The sequences generated in this study will provide valuable resources to better understand the molecular mechanisms and pathways underlying the plethora of secondary metabolites synthesized in *S. hispanica* leaves and would contribute to future research aimed at further improvement of these characteristics.

2. Materials & Methods

2.1. Plant Materials

Seeds of *S. hispanica* Pinta cultivar were germinated in Sunshine[®] All-Purpose potting mix and maintained in Conviron[®] growth chamber at 22 °C with a 16-h light and 8-h dark photoperiod and 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ light intensity for two weeks. At the four-leaf developmental stage, a pair of newly emerged leaves were harvested at Zeitgeber Time four (ZT4; four hours after lights on) and immediately frozen in liquid nitrogen. Roots were washed thoroughly with tap water before harvesting. Tissues from six seedlings were combined together as one biological replicate. A total of three biological replicates were collected.

2.2. RNA Extraction, Library Construction and Illumina Sequencing

RNA was extracted from leaf and root tissues using TRIzol[®] (Invitrogen) according to manufacturer's instructions. RNA was further purified using Spectrum[™] Plant Total RNA Kit (Sigma-Aldrich) and subjected to on-column DNase treatment. RNA quality and quantity were assessed using Agilent 2100 Bioanalyzer (Agilent Technologies). Then, 500 ng total RNA was used for RNA-seq library preparation according to the protocol described by [24]. Briefly, mRNA was isolated using oligo(dT) coated magnetic beads (Invitrogen) and treated with DNase followed by first and second strand cDNA synthesis. The cDNA was fragmented using divalent cations and enriched for fragments around 300 bp. Finally, custom barcoded adaptors were ligated to the fragments followed by 10 cycles of PCR enrichment of the library products. The barcoded libraries were pooled together and subjected to 150 bp paired-end sequencing on an Illumina HiSeq4000 machine (UC Berkeley; Vincent J. Coates Genomics Sequencing Laboratory).

2.3. Bioinformatic Analysis

From the raw sequences, the adaptors and low-quality bases were trimmed using Trimmomatic version 0.36 with 100 bp minimum length cutoff [25]. The remaining high-quality reads were *de novo* assembled using Trinity [26] version 2.5.1. The assembled contigs were clustered using the CD-HIT-EST program with a 90% identity threshold [27] and the longest representative sequence in each cluster was selected using a custom python script. The completeness of the assembly was evaluated by Benchmarking Universal Single-Copy Orthologs (BUSCO) [28] using the embryophyta_odb9 database containing 1440 categories. The contigs were annotated using the uniprot database, in addition to Arabidopsis and tomato protein sequences using DIAMOND [29] version 0.9.22. Gene Ontology (GO) annotation was performed using AgBase version 2.0 [30] and GO enrichment analysis was conducted using PANTHER version 11 with conservative Bonferroni correction for multiple testing [31]. The RNA-seq reads were mapped against the *de novo* transcriptome assembly using Salmon version 0.8.1 [32] and differential gene expression analysis was performed using the generalized linear model

(glm) functionality of the edgeR package [33]. Contigs with at least two-fold expression difference between leaf and root and False Discovery Rate (FDR) < 0.01 were considered differentially expressed.

2.4. Clustering

The differentially expressed contigs (DECs) were hierarchically clustered into 30 groups by expression similarity using the hclust function of the stats package (R Core Team, 2018) version 3.6.0. The clustering was carried out using the complete method, which considers the largest value of dissimilarities between clusters. The package dendextend version 1.9.0 [34] was used to plot a dendrogram demonstrating members which are similar in a subgroup, and members which are dissimilar and in distinct clusters. The results were then put through log transformation and displayed with a heatmap, using the gplots package version 3.0.1 [35].

2.5. Phylogenetic Analysis

The phylogenetic relationship among 37 plant species representing seven families was assessed using the chloroplast Maturase K (matK) gene. The protein sequences of the MatK gene were downloaded from the Genbank non redundant protein database. The protein sequences of tomato Terpene synthase genes were obtained from [16] and blasted against the assembled *S. hispanica* transcriptome to identify putative terpene synthase family members. Among the matching contigs, further filtering was done based on the presence of Terpene synthase family, metal binding domain (pfam03936). The sequences were aligned using the ClustalW program and the phylogenetic tree was constructed using Phylogeny.fr [36] with the maximum likelihood method and 1000 bootstrap replicates.

2.6. Hormone and Stress Treatments

Seeds of *S. hispanica* Pinta cultivar were germinated and maintained as described above in 16 h light/8 h dark photoperiod. Two-week-old seedlings were sprayed with 50 µM Gibberellic acid (GA), 100 µM indole-3-acetic acid (IAA), 100 µM abscisic acid (ABA) in 0.05% tween 20 solution. Control plants were sprayed with the same amount of 0.05% tween 20 solution. Seedlings were also exposed to heat (37 °C) and cold (4 °C) treatments. The GA, ABA, cold, and heat treatments were performed at ZT4 (four hours after lights on) while the IAA treatment was done at ZT12 (12 hours after lights on). Leaves were harvested three hours after hormone treatments and one hour after cold and heat treatments and were immediately frozen in liquid nitrogen.

2.7. cDNA Synthesis & qPCR Analysis

Total RNA was extracted from frozen leaf and root samples and DNase treated as described above. cDNA was prepared from 100 ng total RNA using Superscript III first strand cDNA synthesis kit (Invitrogen USA). qPCR primers were designed using the online Primer 3 software (Additional file 1). The housekeeping genes Serine/threonine-protein phosphatase 2A (PP2A) and Cyclophilin (CYP) were used as internal controls to normalize the data [37]. Three biological replicates were used. qPCR was run on the Bio-Rad CFX96 machine using the following conditions: 95 °C for 5 min, followed by 40 cycles of 95 °C for 20 sec and 60 °C for 1 min. The fold change in gene expression levels was calculated using the $2(-\Delta\Delta CT)$ method [38]. Significant differences in gene expression levels were determined using a *t*-test.

3. Results and Discussion

3.1. This Sequencing and de novo Assembly

To obtain an overview of the *S. hispanica* transcriptome, RNA-Seq libraries were prepared from leaf and root tissues of two-week-old seedlings. A total of 90 million high quality 150 bp paired-end reads were generated. The reads were *de novo* assembled into 279,905 contigs greater than 300 bp, which is considerably higher than the number of protein-coding genes in well studied plants with

similar size genomes such as *Arabidopsis* (35,386), *Medicago truncatula* (62,319), *Ananas comosus* (27,024), and *Populus trichocarpa* (73,013) (<https://phytozome.jgi.doe.gov>). Unlike genome-guided assemblers, the currently available *de novo* assembly programs are known to generate a high level of redundancy. Among the contributors of this redundancy are the sequencing errors and single nucleotide polymorphisms (SNPs), which create mismatches [39]. Accordingly, redundant sequences get generated as the assembly programs fail to consolidate highly similar sequences. This fact is exacerbated with increasing the number of reads used in the transcriptome assembly [39]. To assess the completeness of our transcriptome and the level of redundancy, BUSCO analysis was performed. Among the 1440 BUSCO groups searched, 4% were “complete and single-copy”, 88.8% were “complete and duplicated”, 2.2% “fragmented”, and the remaining 5% were “missing” (Figure 1a). Accordingly, the completeness score was 92.8%. This indicates that most of the evolutionarily conserved core plant gene set is present in our assembly, suggesting a high-quality assembly. However, as anticipated, high level (88.8%) redundancy was detected. The redundant sequences in our initial assembly were consolidated using the CD-HIT-EST program, which resulted in 103,367 contigs and BUSCO output of 50.2% “complete and single-copy”, 42% “complete and duplicated”, 2.4% “fragmented”, and 5.4% “missing” while maintaining completeness score of 92.2% (Figure 1a). The remaining redundancy could be attributed to the heterogeneity of the *S. hispanica* genotype sequenced in this study, in addition to possible sequencing and assembly errors. Around 40% of the assembled contigs had a length distribution between 300 and 1000 base pairs (bp) (Figure 1b), with N50 equal to 2330 bp and a maximum transcript length of 26,500 bp.

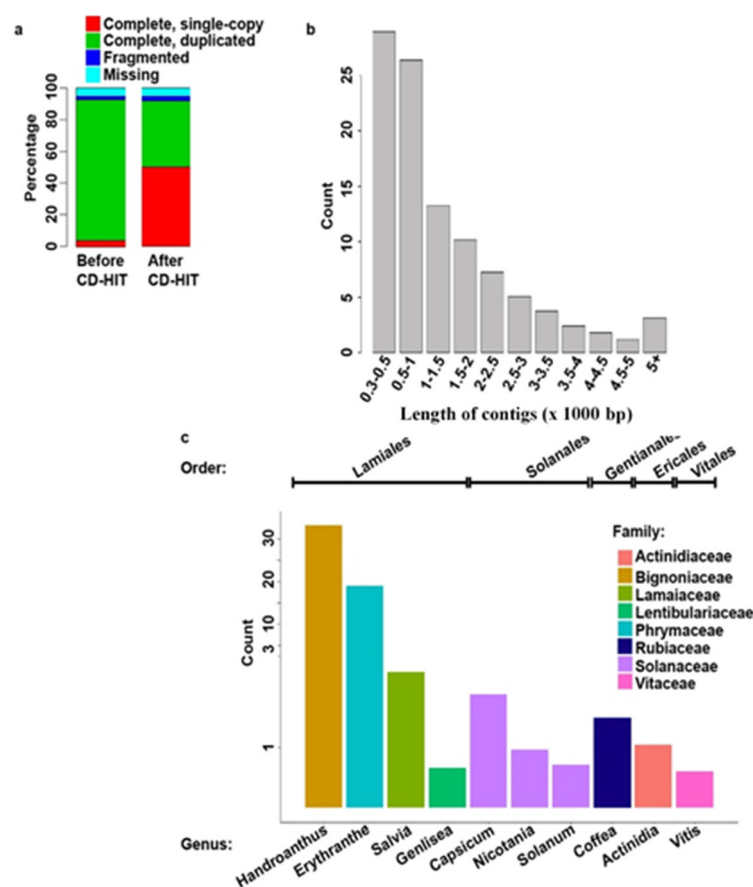


Figure 1. Statistics of the transcriptome assembly. (a) BUSCO results showing the transcriptome completeness and duplication level both before and after consolidating the redundant sequences using CD-HIT-EST; (b) length distribution of the assembled transcripts in basepair (bp); (c) the number of *S. hispanica* sequences matching to the top ten genera belonging to eight plant families. The x-axis shows the top ten genera; the y-axis shows the number of genes ($\times 1000$).

3.3. Differential Gene Expression and GO Enrichment Analysis

Gene expression profiles vary considerably among the different tissues and organs, giving each its unique characteristics. To identify tissue specific transcriptome profiles, the leaf and root RNA-Seq reads were independently mapped against the assembled contigs and differential expression analysis was performed using the EdgeR package [33]. A total of 14,267 contigs showed a significant difference (fold change ≥ 2 ; FDR < 0.01) in expression, among which 6151 and 8116 contigs were upregulated in the leaf and root, respectively (Table S2). Enrichment analysis of the differentially expressed contigs (DECs) and comparison between leaf and root tissues identified diverse and non-overlapping GO terms (Figure 3; Table S3). Hierarchical clustering of the DECs into 30 clusters according to their expression levels also showed both tissue specific and general gene expression patterns (Figure 4a; Table S4). For example, the root specific cluster 28 was enriched for lignin metabolic process and defense response (Figure 4b), while the leaf specific cluster 16 was enriched for photosynthesis (Figure 4c). The details of the clustering results are provided in Table S4 and could be used by the scientific community to predict possible functions of unknown genes.

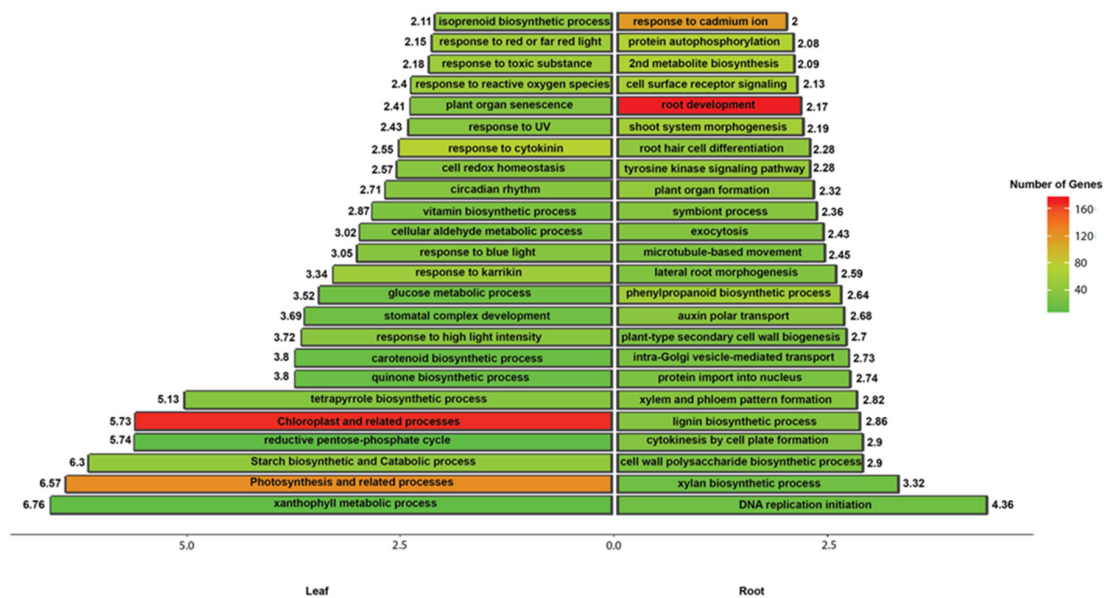


Figure 3. Gene Ontology (GO) enrichment analysis of the differentially expressed contigs (DECs) identified in leaf and root tissues. The left bars and the right bars show the GO terms enriched within the Biological Process ontologies in the leaf and root, respectively. The x-axis shows the fold change of each GO term, which is also provided as a number next to each bar (GO term). The colors represent the number of contigs represented within each GO term (bar).

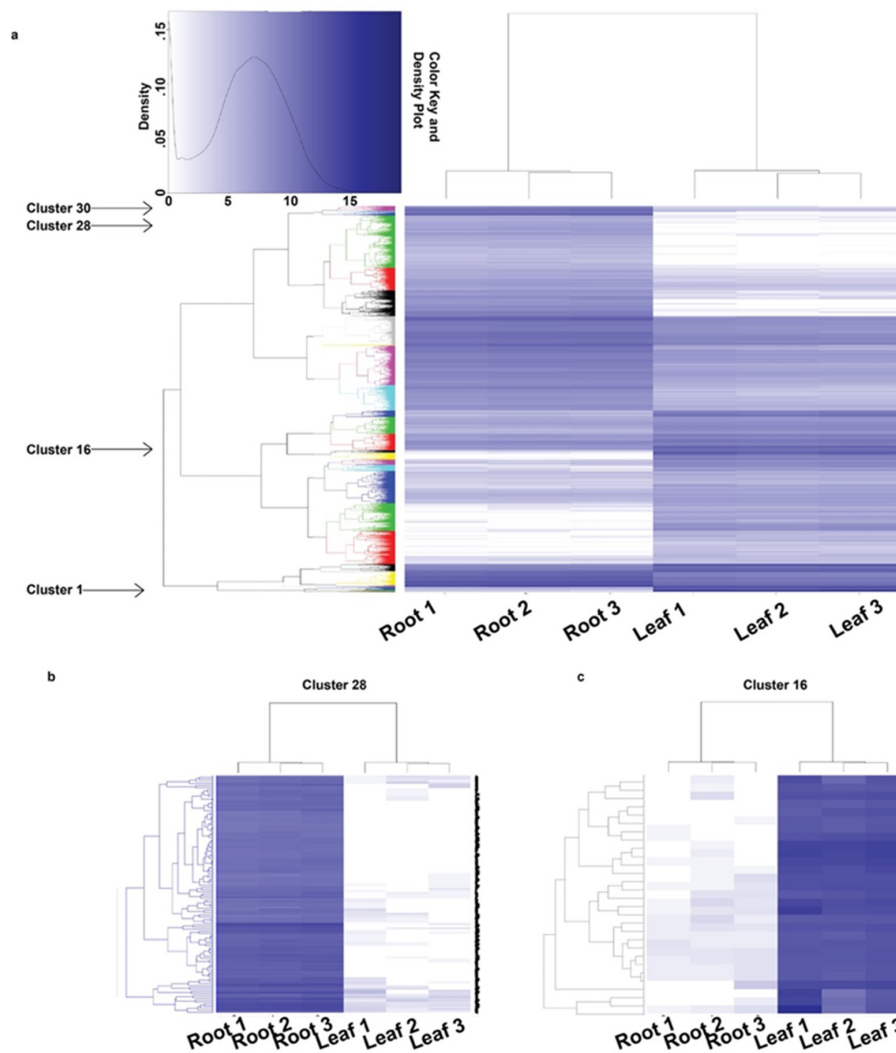


Figure 4. Hierarchical clustering and expression patterns of differentially expressed contigs (DECs) in root and leaf tissues. (a) Hierarchical clustering of the DECs into 30 clusters; (b) magnification of cluster 28; (c) magnification of cluster 16. Root 1,2,3 represent the three root biological replicates used in the analysis and Leaf 1,2,3 represent the three leaf biological replicates used in the analysis. Each horizontal line represents a contig among the 14,267 contigs showing significant difference (fold change ≥ 2 ; FDR < 0.01) in expression. Colors depict expression levels, with darker colors corresponding to higher expression.

3.4. Enrichment of Vitamin Biosynthetic Genes and RT-qPCR Validation of Gene Expression

On dry weight basis, chia seed contain 8.83 mg niacin, 0.17 mg riboflavin (Vitamin B2), 0.62 mg thiamin (Vitamin B1), and 4.3mg vitamin A per 100 g of seed. Thus, from a nutritional standpoint, chia is a good source of B vitamins [41]. In comparison with rice and maize seeds, chia seeds contain more niacin and comparable amounts of thiamine and riboflavin [42]. Riboflavin plays a role in induction of plant defense responses [43]. Similarly, thiamine is an essential cofactor for a number of important metabolic pathways [44] and its deficiency is surprisingly common in humans which causes neurological and cardiovascular problems, weight loss, and confusion [45]. While cereal grains represent a good source of thiamine (0.55 mg/100 g of whole-wheat flour), most is lost during processing (0.06 mg/100 g white flour) [45]. Chia seed, which is eaten raw, contains 0.62 mg/100 g thiamine, representing a valuable source.

In this study, the vitamin biosynthetic process GO term was significantly enriched within the 6151 contigs upregulated in the *S. hispanica* leaf compared to the root, and is represented by riboflavin,

thiamine, pyridoxine (Vitamin B6), ubiquinone, and other terpenoid-quinone metabolic pathways. According to our comparative analysis, the Arabidopsis homologs of the majority of the biosynthetic genes functioning within these pathways were identified in our transcriptome and a number of them were full length sequences (Table S5). RT-qPCR quantification of the riboflavin biosynthesis protein (RIBA2) homolog and homologs of two Arabidopsis thiamine biosynthetic genes were performed and shown to be upregulated in the leaf compared to the root (Figure 5), consistent with the RNA-Seq results.

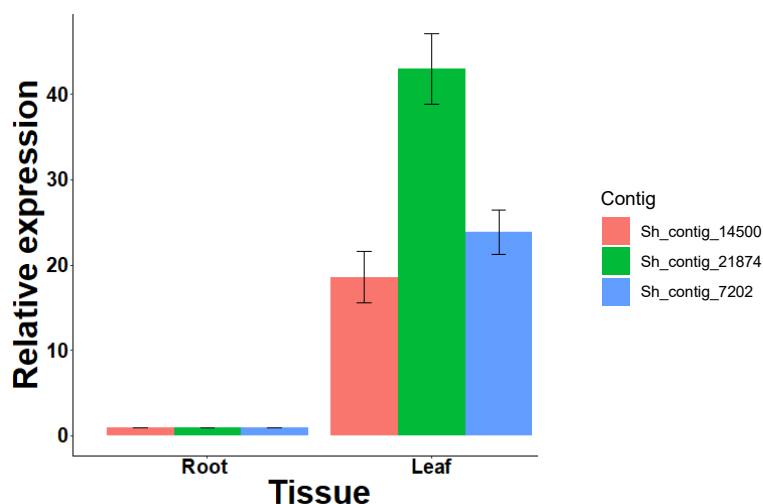


Figure 5. Relative expression analysis of putative riboflavin and thiamine biosynthetic genes in *S. hispanica* leaf and root tissues. Error bars represent the standard error of the mean (SEM). Asterisks indicate significant differences compared to the controls ($P < 0.05$; Student's *t*-test).

3.5. *S. hispanica* Terpene Synthase Family Genes

Terpenes (monoterpenes, sesquiterpenes and diterpenes) are secondary plant metabolites that play an important role in multiple biological functions [46]. Volatile terpenoids, mainly represented by isoprene (C5), monoterpenes (C10) and sesquiterpenes (C15), constitute the largest class of plant volatile compounds [47]. (E)- β -caryophyllene is a type of sesquiterpene that has been identified in a number of plant essential oils such as oregano (*Origanum vulgare* L.), cinnamon (*Cinnamomum* spp.), black pepper (*Piper nigrum* L.), and chia leaves [7,9,48–50]. (E)- β -caryophyllene has been shown to selectively bind to the THC binding site in the CB2 receptor, leading to cellular activation and anti-inflammatory effects [51]. Several other biological activities are attributed to β -caryophyllene such as antibiotic, antioxidant, anticarcinogenic and local anesthetic activities [52]. The plant Terpene synthases (TPSs) are the enzymes responsible for the formation of these diverse terpene metabolites [53]. The majority of the sesquiterpenes found in the tomato leaf trichomes have been attributed to the activities of TPS9 (Sst1), encoding a germacrene C synthase, and TPS12 (CAHS), encoding β -caryophyllene and α -humulene synthase [16,54,55]. In our assembled transcriptome, we identified 30 contigs belonging to the TPS family, evidenced by the presence of the conserved Terpene synthase family metal binding domain (pfam03936). According to a phylogenetic analysis using the 29 functional tomato TPS genes as reference, the 30 *S. hispanica* putative TPS sequences clustered with TS-a, TS-b, TS-c, TPS-g and TS-e/f groups (Figure 6). Sh_contig_19087 and Sh_contig_14884 clustered with tomato TPS12, which encodes for (E)- β -caryophyllene/ α -humulene synthase [16,54]. Sh_contig_56017, Sh_contig_6900, and Sh_contig_18501 clustered with tomato TPS20, which encodes β -phellandrene synthase [16,56]. Our analysis identified four *S. hispanica* Terpene synthases that belong to TPS-g. In tomato, this clade encodes two enzymes with linalool/nerolidol synthase activity [16]. The expression of three contigs within TPS-a, two contigs within TPS-b, and two contigs within TPS-g were investigated following hormone, heat, and cold treatments. Plant hormones have a pivotal role in many physiological processes including development, immunity, adaptation to the environment, and primary metabolism. Similarly,

hormones have been shown to modulate a plant's secondary metabolism, including its terpenoid levels. The total amount of sweet basil (*Ocimum basilicum*) terpenoids significantly increased after Methyl Jasmonate (MeJA) treatment [57,58], and cotton plants treated with MeJA emitted elevated levels of volatile terpenes [59]. In field-grown grapevines, exogenous ABA application significantly increased Sesquiterpene Nerolidol production [60]. In *Panax quinquefolium* hairy root cultures, ABA induced the synthesis of triterpenoid saponins named ginsenosides [61]. Tanshinone production in the *Salvia miltiorrhiza* hairy root system increased upon 3.76 μ M ABA treatment [62]. Exogenous GA3 increased the percentage of some sesquiterpenes (Caryophyllene, Spathulenol, β -Eudesmol, α -Bisabolol) and decreased the percentage of tree sesquiterpenes (β -Farnesene, α -Humulene, Germacrene D) [63].

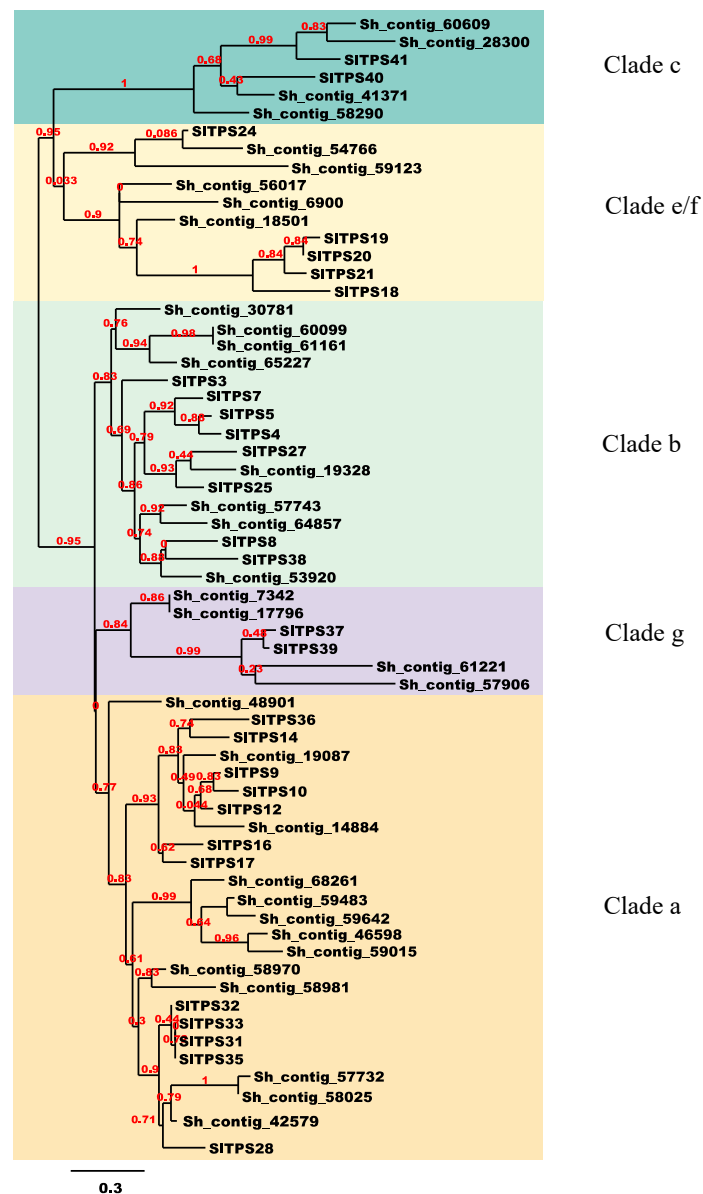


Figure 6. Phylogenetic analysis of *S. hispanica* Terpene synthase proteins. Maximum-likelihood phylogenetic tree of putative *Salvia hispanica* and *Solanum lycopersicum* Terpene synthase proteins. Bootstrap values for 100 replicates are indicated in red. “Sh” and “SI” refer to *Salvia hispanica* and *Solanum lycopersicum*, respectively.

Hormone and abiotic stress treatments also effected the expression of some TPS gene family members in *S. hispanica*. The gene expression of all the three putative TPS-a members tested in this study were significantly repressed following ABA treatment (Figure 7a–c); this also occurred in Sh_contig_64857, which is a member of TPS-b (Figure 7f). In contrast, the expression of Sh_contig_61221 (TPS-g member) significantly increased following ABA treatment (Figure 7e). Interestingly, IAA and GA had no effect on the expression of the TPS genes tested in this study. Heat stress had mixed effects on the different group members. The relative expression of Sh_contig_42579 (TPS-a member) was significantly reduced, while that of Sh_contig_65227 (TPS-b member) was significantly increased three hours after heat treatment at 37 °C. Upon cold stress treatment, only the expression of Sh_contig_61221 (TPS-g member) was significantly induced. While gene expression analysis provides some ideas regarding the possible roles of the TPS genes in various plant physiological and adaptation processes, more direct characterization of their functions requires the expression of these genes in *Escherichia coli* followed by enzymatic assays.

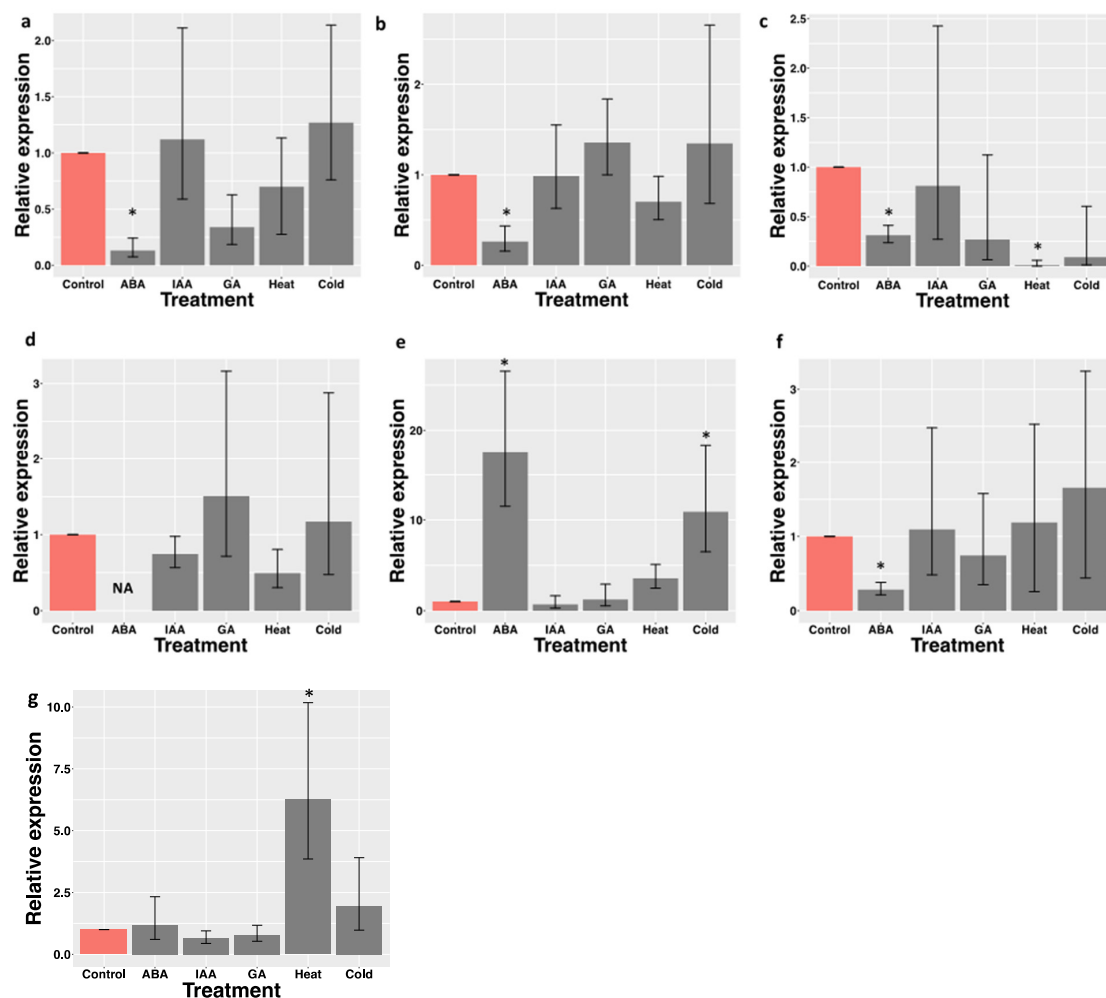


Figure 7. Relative expression analysis of putative terpene biosynthetic genes in *S. hispanica* leaves exposed to abiotic stress (heat and cold) and hormone treatments. ABA: abscisic acid; IAA: indole acetic acid; GA: gibberellic acid. (a) Sh_contig_19087; (b) Sh_contig_46958; (c) Sh_contig_57906; (d) Sh_contig_42579; (e) Sh_contig_61221; (f) Sh_contig_64857 (g) Sh_contig_65227. Error bars represent the standard error of the mean (SEM). Asterisks indicate significant differences ($P < 0.05$; Student's *t*-test).

4. Conclusions

Recent advances in sequencing technologies have significantly contributed to our understanding of complex biological processes, especially in non-model plant species. Here, we sequenced and assembled the tissue specific transcriptome of *S. hispanica*, and unlike previous transcriptome studies in this species, we made all the sequences available to the plant science community as supplementary information. The generated data and downstream analysis in terms of differential gene expression, clustering, phylogenetic, and RT-qPCR analysis, is a first step to better understand the various beneficial characteristics of this plant species at the molecular level. In addition, future comparative analysis of the *S. hispanica* sequences with closely related species with sequenced genomes would identify lineage-specific genes and further elucidate specific molecular pathways.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2223-7747/9/3/405/s1>, Figure S1: Phylogenetic analysis of *S. hispanica*, File S1: Sequences of *S. hispanica* transcriptome assembled in this study, Table S1: Annotation of the assembled contigs, Table S2: The counts of the differentially expressed genes, Table S3: Enrichment analysis within individual clusters, Table S4: Gene names and annotations representing individual clusters, Table S5: *S. hispanica* homologs of Thiamine and Riboflavin biosynthetic genes.

Author Contributions: Conceptualization, H.S.A.; Methodology, H.S.A. and J.C.; Software, J.W. and H.S.A.; Formal Analysis, H.S.A., J.W., and J.C.; Resources, H.S.A.; Data Curation, H.S.A. and J.W.; Writing—Original Draft Preparation, H.S.A. and J.C.; Writing—Review & Editing, H.S.A.; Visualization, J.W.; Supervision, H.S.A.; Project Administration, H.S.A.; Funding Acquisition, H.S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by USDA-NIFA (grant no. 2018-06277) to H.S.A.

Acknowledgments: We thank Ehsan Yaghmaei for help with the analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ayerza, R.; Coates, W. *Chia: Rediscovering a Forgotten Crop of the Aztecs*; University of Arizona Press: Tucson, AZ, USA, 2005.
2. Cahill, J.P. Ethnobotany of Chia, *Salvia hispanica* L. (Lamiaceae). *Econ. Bot.* **2003**, *57*, 604–618. [[CrossRef](#)]
3. Reyes-Caudillo, E.; Tecante, A.; Valdivia-Lopez, M.A. Dietary fibre content and antioxidant activity of phenolic compounds present in Mexican Chia (*Salvia hispanica* L.) seeds. *Food Chem.* **2008**, *107*, 656–663. [[CrossRef](#)]
4. Sandoval-Oliveros, M.R.; Paredes-Loópez, O. Isolation and characterization of proteins from Chia seeds (*Salvia hispanica* L.). *J. Agric. Food Chem.* **2013**, *61*, 193–201. [[CrossRef](#)] [[PubMed](#)]
5. Ayerza, R.; Coates, W. Protein content, oil content and fatty acid profiles as potential criteria to determine the origin of commercially grown chia (*Salvia hispanica* L.). *Ind. Crops Prod.* **2011**, *34*, 1366–1371. [[CrossRef](#)]
6. Ouzounidou, G.; Skiada, V.; Papadopoulou, K.K.; Stamatis, N.; Kavvadias, V.; Eleftheriadis, E.; Gaitis, F. Effects of soil pH and arbuscular mycorrhiza (AM) inoculation on growth and chemical composition of chia (*Salvia hispanica* L.) leaves. *Braz. J. Bot.* **2015**, *38*, 487–495. [[CrossRef](#)]
7. Elshafie, H.S.; Aliberti, L.; Amato, M.; De Feo, V.; Camele, I. Chemical composition and antimicrobial activity of chia (*Salvia hispanica* L.) essential oil. *Eur. Food Res. Technol.* **2018**, *244*, 1675–1682. [[CrossRef](#)]
8. Pascual-Villalobos, M.; Correal, E.; Molina, E.; Martínez, J. *Evaluación y Selección de Especies Vegetales Productoras de Compuestos Naturales con Actividad Insecticida*; Centro de Investigación y Desarrollo Agroalimentario (CIDA): Murcia, Spain, 1997.
9. Ahmed, M.; Ting, I.P.; Scora, R.W. Leaf oil composition of *Salvia hispanica* L. from three geographical areas. *J. Essent. Oil Res.* **1994**, *6*, 223–228. [[CrossRef](#)]
10. Cheng, A.X.; Lou, Y.G.; Mao, Y.B.; Lu, S.; Wang, L.J.; Chen, X.Y. Plant terpenoids: Biosynthesis and ecological functions. *J. Integr. Plant Biol.* **2007**, *49*, 179–186. [[CrossRef](#)]
11. Dudareva, N.; Pichersky, E.; Gershenzon, J. Biochemistry of plant volatiles. *Plant Physiol.* **2004**, *135*, 1893–1902. [[CrossRef](#)]

12. Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **2011**, *66*, 212–229. [[CrossRef](#)]
13. Mumm, R.; Posthumus, M.A.; Dicke, M. Significance of terpenoids in induced indirect plant defence against herbivorous arthropods. *Plant Cell Environ.* **2008**, *31*, 575–585. [[CrossRef](#)]
14. Pichersky, E.; Gershenzon, J. The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. *Curr. Opin. Plant Biol.* **2002**, *5*, 237–243. [[CrossRef](#)]
15. Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4126–4133. [[CrossRef](#)] [[PubMed](#)]
16. Falara, V.; Akhtar, T.A.; Nguyen, T.T.H.; Spyropoulou, E.A.; Bleeker, P.M.; Schauvinhold, I.; Matsuba, Y.; Bonini, M.E.; Schillmiller, A.L.; Last, R.L.; et al. The tomato terpene synthase gene family. *Plant Physiol.* **2011**, *157*, 770–789. [[CrossRef](#)] [[PubMed](#)]
17. Martin, D.M.; Aubourg, S.; Schouwey, M.B.; Daviet, L.; Schalk, M.; Toub, O.; Lund, S.T.; Bohlmann, J. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biol.* **2010**, *10*, 226. [[CrossRef](#)]
18. Huang, M.; Abel, C.; Sohrabi, R.; Petri, J.; Haupt, I.; Cosimano, J.; Gershenzon, J.; Tholl, D. Variation of herbivore-induced volatile terpenes among arabidopsis ecotypes depends on allelic differences and subcellular targeting of two terpene synthases, TPS02 and TPS03. *Plant Physiol.* **2010**, *153*, 1293–1310. [[CrossRef](#)]
19. Aubourg, S.; Lecharny, A.; Bohlmann, J. Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genom.* **2002**, *267*, 730–745. [[CrossRef](#)]
20. Miller, D.J.; Allemann, R.K. Sesquiterpene synthases: Passive catalysts or active players? *Nat. Prod. Rep.* **2011**, *29*, 60–71. [[CrossRef](#)]
21. Schnee, C.; Kollner, T.G.; Held, M.; Turlings, T.C.; Gershenzon, J.; Degenhardt, J. The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1129–1134. [[CrossRef](#)]
22. Lu, X.; Tang, K.; Li, P. Plant metabolic engineering strategies for the production of pharmaceutical terpenoids. *Front. Plant Sci.* **2016**, *7*, 1647. [[CrossRef](#)]
23. Amato, M.; Caruso, M.C.; Guzzo, F.; Galgano, F.; Commisso, M.; Bochicchio, R.; Labella, R.; Favati, F. Nutritional quality of seeds and leaf metabolites of Chia (*Salvia hispanica* L.) from Southern Italy. *Eur. Food Res. Technol.* **2015**, *241*, 615–625. [[CrossRef](#)]
24. Kumar, R.; Ichihashi, Y.; Kimura, S.; Chitwood, D.H.; Headland, L.R.; Peng, J.; Maloof, J.N.; Sinha, N.R. A High-Throughput Method for Illumina RNA-Seq Library Preparation. *Front. Plant Sci.* **2012**, *3*, 202. [[CrossRef](#)] [[PubMed](#)]
25. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
26. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)]
27. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [[CrossRef](#)]
28. Simao, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
29. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [[CrossRef](#)]
30. McCarthy, F.M.; Wang, N.; Magee, G.B.; Nanduri, B.; Lawrence, M.L.; Camon, E.B.; Barrell, D.G.; Hill, D.P.; Dolan, M.E.; Williams, W.P.; et al. AgBase: A functional genomics resource for agriculture. *BMC Genom.* **2006**, *7*, 229. [[CrossRef](#)]
31. Mi, H.; Huang, X.; Muruganujan, A.; Tang, H.; Mills, C.; Kang, D.; Thomas, P.D. PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **2017**, *45*, D183–D189. [[CrossRef](#)]

32. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14*, 417–419. [[CrossRef](#)]
33. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]
34. Galili, T. Dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **2015**, *31*, 3718–3720. [[CrossRef](#)] [[PubMed](#)]
35. Warnes, G.R. Gplots: Various R Programming Tools for Plotting Data. Available online: <http://cranr-project.org/web/packages/gplots/index.html> (accessed on 15 March 2018).
36. Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J.F.; Guindon, S.; Lefort, V.; Lescot, M.; et al. Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **2008**, *36*, W465–W469. [[CrossRef](#)] [[PubMed](#)]
37. Gopalam, R.; Rupwate, S.D.; Tumaney, A.W. Selection and validation of appropriate reference genes for quantitative real-time PCR analysis in *Salvia hispanica*. *PLoS ONE* **2017**, *12*, e0186978. [[CrossRef](#)] [[PubMed](#)]
38. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)] [[PubMed](#)]
39. Huang, X.; Chen, X.G.; Armbruster, P.A. Comparative performance of transcriptome assembly methods for non-model organisms. *BMC Genom.* **2016**, *17*, 523. [[CrossRef](#)]
40. Dong, W.; Liu, J.; Yu, J.; Wang, L.; Zhou, S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **2012**, *7*, e35071. [[CrossRef](#)]
41. Bushway, A.A.; Belyea, P.R.; Bushway, R.J. Chia seed as a source of oil, polysaccharide, and protein. *J. Food Sci.* **1981**, *46*, 1349–1351. [[CrossRef](#)]
42. Marcinek, K.; Krejpcio, Z. Chia seeds (*Salvia hispanica*): Health promoting properties and therapeutic applications—A review. *Rocz. Panstw. Zakl. Hig.* **2017**, *68*, 123–129.
43. Dong, H.; Beer, S.V. Riboflavin induces disease resistance in plants by activating a novel signal transduction pathway. *Phytopathology* **2000**, *90*, 801–811. [[CrossRef](#)]
44. Goyer, A. Thiamine in plants: Aspects of its metabolism and functions. *Phytochemistry* **2010**, *71*, 1615–1624. [[CrossRef](#)] [[PubMed](#)]
45. Fitzpatrick, T.B.; Basset, G.J.; Borel, P.; Carrari, F.; DellaPenna, D.; Fraser, P.D.; Hellmann, H.; Osorio, S.; Rothan, C.; Valpuesta, V.; et al. Vitamin deficiencies in humans: Can plant science help? *Plant Cell* **2012**, *24*, 395–414. [[CrossRef](#)] [[PubMed](#)]
46. Gao, F.; Liu, B.; Li, M.; Gao, X.; Fang, Q.; Liu, C.; Ding, H.; Wang, L.; Gao, X. Identification and characterization of terpene synthase genes accounting for volatile terpene emissions in flowers of *Freesia x hybrida*. *J. Exp. Bot.* **2018**, *69*, 4249–4265. [[CrossRef](#)] [[PubMed](#)]
47. Nagegowda, D.A. Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Lett.* **2010**, *584*, 2965–2973. [[CrossRef](#)]
48. Jayaprakasha, G.K.; Jagan Mohan Rao, L.; Sakariah, K.K. Volatile constituents from *Cinnamomum zeylanicum* fruit stalks and their antioxidant activities. *J. Agric. Food Chem.* **2003**, *51*, 4344–4348. [[CrossRef](#)]
49. Mockute, D.; Bernotiene, G.; Judzentiene, A. The essential oil of *Origanum vulgare* L. ssp. *vulgare* growing wild in vilnius district (Lithuania). *Phytochemistry* **2001**, *57*, 65–69. [[CrossRef](#)]
50. Orav, A.; Stulova, I.; Kailas, T.; Muurisepp, M. Effect of storage on the essential oil composition of *Piper nigrum* L. fruits of different ripening states. *J. Agric. Food Chem.* **2004**, *52*, 2582–2586. [[CrossRef](#)]
51. Gertsch, J.; Leonti, M.; Raduner, S.; Racz, I.; Chen, J.Z.; Xie, X.Q.; Altmann, K.H.; Karsak, M.; Zimmer, A. Beta-caryophyllene is a dietary cannabinoid. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9099–9104. [[CrossRef](#)]
52. Legault, J.; Pichette, A. Potentiating effect of beta-caryophyllene on anticancer activity of alpha-humulene, isocaryophyllene and paclitaxel. *J. Pharm. Pharmacol.* **2007**, *59*, 1643–1647. [[CrossRef](#)]
53. Tholl, D. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* **2006**, *9*, 297–304. [[CrossRef](#)]
54. Schillmiller, A.L.; Miner, D.P.; Larson, M.; McDowell, E.; Gang, D.R.; Wilkerson, C.; Last, R.L. Studies of a biochemical factory: Tomato trichome deep expressed sequence tag sequencing and proteomics. *Plant Physiol.* **2010**, *153*, 1212–1223. [[CrossRef](#)] [[PubMed](#)]
55. Colby, S.M.; Crock, J.; Dowdle-Rizzo, B.; Lemaux, P.G.; Croteau, R. Germacrene C synthase from *Lycopersicon esculentum* cv. VFNT cherry tomato: cDNA isolation, characterization, and bacterial expression of the multiple product sesquiterpene cyclase. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2216–2221. [[CrossRef](#)] [[PubMed](#)]

56. Schillmiller, A.L.; Schauvinhold, I.; Larson, M.; Xu, R.; Charbonneau, A.L.; Schmidt, A.; Wilkerson, C.; Last, R.L.; Pichersky, E. Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10865–10870. [[CrossRef](#)] [[PubMed](#)]
57. Kim, H.J.; Chen, F.; Wang, X.; Rajapakse, N.C. Effect of methyl jasmonate on secondary metabolites of sweet basil (*Ocimum basilicum* L.). *J. Agric. Food Chem.* **2006**, *54*, 2327–2332. [[CrossRef](#)]
58. Li, Z.; Wang, X.; Chen, F.; Kim, H.J. Chemical changes and overexpressed genes in sweet basil (*Ocimum basilicum* L.) upon methyl jasmonate treatment. *J. Agric. Food Chem.* **2007**, *55*, 706–713. [[CrossRef](#)]
59. Rodriguez-Saona, C.; Crafts-Brandner, S.J.; Pare, P.W.; Henneberry, T.J. Exogenous methyl jasmonate induces volatile emissions in cotton plants. *J. Chem. Ecol.* **2001**, *27*, 679–695. [[CrossRef](#)]
60. Alonso, R.; Berli, F.J.; Bottini, R.; Piccoli, P. Acclimation mechanisms elicited by sprayed abscisic acid, solar UV-B and water deficit in leaf tissues of field-grown grapevines. *Plant Physiol. Biochem.* **2015**, *91*, 56–60. [[CrossRef](#)]
61. Kochan, E.; Balcerzak, E.; Szymczyk, P.; Sienkiewicz, M.; Zielinska-Blizniewska, H.; Szymanska, G. Abscisic acid regulates the 3-Hydroxy-3-methylglutaryl CoA Reductase gene promoter and ginsenoside production in *Panax quinquefolium* hairy root cultures. *Int. J. Mol. Sci.* **2019**, *20*, 1310. [[CrossRef](#)]
62. Gupta, S.K.; Liu, R.; Liaw, S.Y.; Chan, H.S.; Tsay, H.S. Enhanced tanshinone production in hairy roots of '*Salvia miltiorrhiza* Bunge' under the influence of plant growth regulators in liquid culture. *Bot. Stud.* **2011**, *52*, 435–443.
63. Mansouri, H.; Asrar, Z.; Amarowicz, R. The response of terpenoids to exogenous gibberellic acid in *Cannabis sativa* L. at vegetative stage. *Acta Physiol. Plant.* **2010**, *33*, 1085–1091. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).