



Chia (*Salvia hispanica*) Gene Expression Atlas Elucidates Dynamic Spatio-Temporal Changes Associated With Plant Growth and Development

Parul Gupta^{1†}, Matthew Geniza^{1,2†}, Sushma Naithani^{1*}, Jeremy L. Phillips¹, Ebaad Haq¹ and Pankaj Jaiswal^{1*}

OPEN ACCESS

¹ Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, United States, ² Molecular and Cellular Biology Graduate Program, Oregon State University, Corvallis, OR, United States

Edited by:

Jill Margaret Farrant,
University of Cape Town, South Africa

Reviewed by:

Atanas Ivanov Atanassov,
Joint Genomic Center, Bulgaria
Ndiko Ndomelele Ludidi,
University of the Western Cape,
South Africa

*Correspondence:

Sushma Naithani
sushma.naithani@oregonstate.edu
Pankaj Jaiswal
pankaj.jaiswal@oregonstate.edu

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Plant Biotechnology,
a section of the journal
Frontiers in Plant Science

Received: 16 February 2021

Accepted: 09 June 2021

Published: 20 July 2021

Citation:

Gupta P, Geniza M, Naithani S,
Phillips JL, Haq E and Jaiswal P
(2021) Chia (*Salvia hispanica*) Gene
Expression Atlas Elucidates Dynamic
Spatio-Temporal Changes Associated
With Plant Growth and Development.
Front. Plant Sci. 12:667678.
doi: 10.3389/fpls.2021.667678

Chia (*Salvia hispanica* L.), now a popular superfood and a pseudocereal, is one of the richest sources of dietary nutrients such as protein, fiber, and polyunsaturated fatty acids (PUFAs). At present, the genomic and genetic information available in the public domain for this crop are scanty, which hinders an understanding of its growth and development and genetic improvement. We report an RNA-sequencing (RNA-Seq)-based comprehensive transcriptome atlas of Chia sampled from 13 tissue types covering vegetative and reproductive growth stages. We used ~355 million high-quality reads of total ~394 million raw reads from transcriptome sequencing to generate *de novo* reference transcriptome assembly and the tissue-specific transcript assemblies. After the quality assessment of the merged assemblies and implementing redundancy reduction methods, 82,663 reference transcripts were identified. About 65,587 of 82,663 transcripts were translated into 99,307 peptides, and we were successful in assigning InterPro annotations to 45,209 peptides and gene ontology (GO) terms to 32,638 peptides. The assembled transcriptome is estimated to have the complete sequence information for ~86% of the genes found in the Chia genome. Furthermore, the analysis of 53,200 differentially expressed transcripts (DETs) revealed their distinct expression patterns in Chia's vegetative and reproductive tissues; tissue-specific networks and developmental stage-specific networks of transcription factors (TFs); and the regulation of the expression of enzyme-coding genes associated with important metabolic pathways. In addition, we identified 2,411 simple sequence repeats (SSRs) as potential genetic markers from the transcripts. Overall, this study provides a comprehensive transcriptome atlas, and SSRs, contributing to building essential genomic resources to support basic research, genome annotation, functional genomics, and molecular breeding of Chia.

Keywords: Chia (*Salvia hispanica* L.), transcriptome, polyunsaturated fatty acid, superfood, expression atlas, pseudocereal grain, omega-3 fatty acids, BUSCO plant gene expression

INTRODUCTION

Salvia hispanica L. (Chia), an annual herbaceous plant of the Lamiaceae (mint) family, is a native of Central America's highlands (Cahill, 2005; Ixtaina et al., 2008; Baginsky et al., 2016). It is a photoperiod-sensitive short-day flowering plant, which usually grows about 1 m in height and produces raceme inflorescence bearing small purple flowers. Chia, which is cultivated primarily for its seeds, is a core ingredient of the Mayan and Aztec population's diet. Recently, its consumption has grown outside South America due to its rich nutritional and gluten-free characteristics (Mohd Ali et al., 2012). Chia seeds contain ~40% oil by weight the majority of which are omega-3 and omega-6 polyunsaturated fatty acids (PUFAs) (Mohd Ali et al., 2012). The seeds are also rich in protein (15–20%), dietary fiber (20–40%), minerals (4–5%), and antioxidants (Reyes-Caudillo et al., 2008; Ayerza and Coates, 2009; Muñoz et al., 2013). These nutritional attributes have made Chia a desirable superfood and a pseudocereal. Several dietary studies in humans and mouse models suggest that a diet supplemented with Chia seeds resulted in improving muscle lipid content, cardiovascular health, total cholesterol ratio, triglyceride content (Vuksan et al., 2007, 2010, 2017a,b; Oliva et al., 2013; Valdivia-López and Tecante, 2015; Ullah et al., 2016; Marcinek and Krejpcio, 2017), and helped to attenuate blood glucose levels in patients with type-2 diabetes (Vuksan et al., 2007; Chicco et al., 2009; Peiretti and Gai, 2009; Oliva et al., 2013). Besides its food and nutrition value, Chia is a rich source of many other useful products. Its leaves contain various essential oils, such as β -caryophyllene, globulol, γ -muroloeno, β -pinene, α -humulene, germacrene, and widdrol, which are known to have insect repellent or insecticidal properties (Amato et al., 2015; Elshafie et al., 2018).

High-throughput gene expression datasets from various oilseed crops, such as soybean (*Glycine max*), peanut (*Arachis hypogaea*), false flax (*Camelina sativa*) (Libault et al., 2010; Severin et al., 2010; Clevenger et al., 2016; Kagale et al., 2016), have been instrumental in understanding the key regulators of metabolic and developmental processes (Druka et al., 2006; Sekhon et al., 2011; Stelpflug et al., 2016; Cañas et al., 2017; Kudapa et al., 2018) and for translational research leading to crop improvement. Despite being a valued crop, the genomics resources available for Chia are limited. A few previous studies focusing on fatty acid metabolism have generated transcriptome data and simple sequence repeats (SSRs) from Chia seeds (Sreedhar et al., 2015; Peláez et al., 2019). Also, Wimberley et al. (2020) reported leaf and root transcriptomes and identified terpenoid biosynthesis genes. To our knowledge, an extensive plant structure and the developmental stage-specific transcriptomes for Chia have not yet been reported.

We took an initiative to build a gene expression atlas for Chia from the 13 plant structures collected from a range of vegetative and reproductive stages. Furthermore, we identified SSRs within the transcribed region of the Chia genome; added functional-structural annotations to transcripts; analyzed the differential expression of transcripts across spatial and temporal scales; and

conducted a pathway enrichment analysis to understand the regulation of Chia metabolism throughout its development.

RESULTS

Sequencing and *de novo* Assembly of Chia Transcriptome

The transcriptome data were generated from the 13 Chia plant structures, including mature dry seeds, early seedling shoots, leaves (representing the developmental stages P1–P7), an internode between P5 and P6 leaves, top and bottom halves of pre-anthesis raceme inflorescence, and flowers from the day of anthesis [Days after flowering (DAF)], and 5 day post-anthesis (Table 1). A 101-basepair (bp) paired-end sequencing of the 39 complementary DNA (cDNA) libraries [three biological replicates for each sample prepared from the poly-A enriched messenger RNA (mRNA)] resulted in 393,645,776 sequence reads and ~80 Gb of the nucleotide sequence (Supplementary Material 1). The high-quality reads were assembled at 65 and 75 k-mer lengths, and then, after merging both k-mer assemblies for each tissue type, their unique transcripts were generated. The number of assembled transcripts ranged from 27,066 to 43,491 for tissue-specific assemblies (Figure 1A). Among the vegetative tissues, D69-P1-P2 leaf samples contained a maximum (43,491), and the seed contained a minimum (27,066) number of assembled transcripts (Figure 1A). Among reproductive structures such as inflorescence and flowers, the maximum number of transcripts (43,418) with an average length of about 1,000 bases was observed in the top half of the D158-raceme inflorescence sample (Figure 1A).

Additionally, the high-quality paired-end reads (352,976,255) from all tissue libraries were pooled and assembled at 67 and 71 k-mer lengths by using Velvet (Zerbino and Birney, 2008) and Oases (Schulz et al., 2012). Chia transcript isoforms generated by each k-mer (67 and 71 k-mer lengths) assembly were merged to represent a total of 145,503 unique transcripts (≥ 201 bases in length) (Figure 1B). The use of the Cluster Database at High Identity with Tolerance-Expressed Sequence Tag (CD-HIT-EST) algorithm (Li and Godzik, 2006) for removing redundant transcripts (displaying $\geq 90\%$ similarity) yielded 82,663 transcripts (Figure 1B). In parallel, we used a quality assessment software, the TransRate (Smith-Unna et al., 2016), on the original assembly that yielded 35,461 transcripts (Figure 1B). TransRate detects the redundant transcripts by aligning the reads to multiple transcripts and assigns all of them to the transcript that best represents a canonical form. We observed that the assembly produced by CD-HIT-EST experienced only a minor loss in the percentage of reads aligned. The assembly produced by TransRate, which utilizes Salmon (Patro et al., 2017) to estimate transcript abundance with map-based methods, contained nearly 50% fewer reads aligned in comparison to the CD-HIT-EST assembly. Furthermore, we used a quality assessment tool, QAST (Mikheenko et al., 2016), on the original assembly and each of the redundancy reduced assemblies (Supplementary Material 2). Both original

Abbreviations: DAF, Days after flowering; DAS, Days after sowing.

TABLE 1 | Description of the plant material used for generating the Chia transcriptome atlas.

Growth stage	Sample collection days after sowing (DAS)	Sample description	Sample name
Vegetative	Day 0	Dry Seed	Seed
	Day 3	Green cotyledon	D3-Cotyledon
	Day 3	Above ground shoot parts (whole shoot)	D3-Shoot
	Day 12	Above ground shoot parts (whole shoot)	D12-Shoot
	Day 12	Very first/youngest leaf at the shoot apex	D12-P1
	Day 69	First and second leaves at the shoot apex	D69-P1-P2
	Day 69	Third and fourth leaves at the shoot apex	D69-P3-P4
	Day 69	Fifth, sixth, and seventh leaves at the shoot apex	D69-P5-P6-P7
	Day 69	Internode between the 5th and 6th leaves	D69-Internode
	Reproductive	Day 158	The top half of the raceme inflorescence (pre-anthesis)
Day 158		The bottom half of the raceme inflorescence (with pre-anthesis flowers)	D158-RacemeBottomHalf
Day159		Flowers from Day-1 of flowering (anthesis stage)	D159-Flowers
Day 164		Flowers from Day-5 of flowering (anthesis stage)	D164-Flowers

and TransRate assemblies had better statistics in transcript number and length, and also contained the worst statistics in the complementing category (**Supplementary Material 2**). The assembly produced by CD-HIT-EST represented the most moderate version. Based on the quality assessment and alignment statistics, we pursued the CD-HIT-EST assembly with 82,663 transcripts for downstream analyses. A workflow for an assembly and a downstream analysis is shown in **Supplementary Material 3**. The Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis suggests that our *de novo* assembled transcriptome contains complete gene sequences for 86.4% of the Chia genes (**Supplementary Material 4**).

Functional Annotation of Chia Transcriptome

We compared the 82,663 assembled Chia transcripts to the publicly available genomes and gene models of eudicots using BLASTx and tBLASTx (Mount, 2007) to estimate approximate coverage of genes represented in the assembled transcriptome. More than 84% (the number that is very close to the BUSCO analysis) of the assembled Chia transcripts were mapped to the closely related *Salvia miltiorrhiza* (Wenping et al., 2011) and *Salvia splendens* (Ge et al., 2014) transcriptomes (**Figure 1C**). The dispersion of coverage within the genus is not surprising since the *Salvia* genus is very diverse. Both *S. miltiorrhiza* and *S. splendens* share a common center of origin in China whereas *S. hispanica* originated in Central America. In the Lamiaceae family, about 56% of the Chia transcripts were mapped to the transcriptomes from the members of the *Mentha* genus, namely water mint (*M. aquatica*), peppermint (*M. piperita*), and spearmint (*M. spicata*) (Ahkami et al., 2015), except that 80% of the Chia transcripts were mapped to *M. longifolia* genes identified in its genome assembly (Vining et al., 2017). Moving up the taxonomic rank to the order of Lamiales, 75% of the Chia transcripts were mapped to sesame (*Sesamum indicum*) (Zhang et al., 2013), an oilseed crop. A total of 71% and 74% Chia transcripts were mapped to a proteome set of the model plants *Arabidopsis thaliana* and

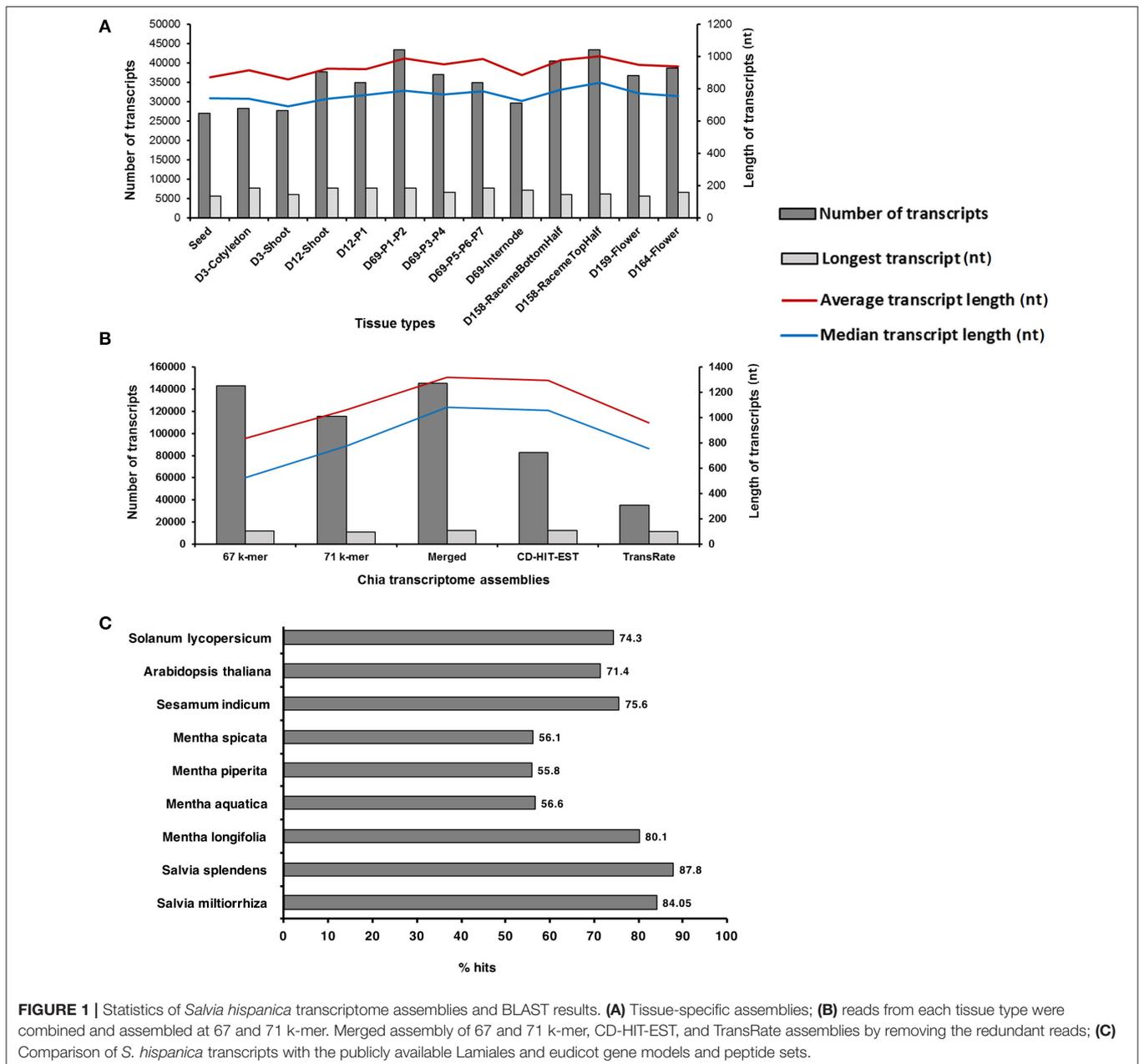
the *Solanum lycopersicum* (tomato), respectively (**Figure 1C**). Although the assembled transcriptomes were not available, two publicly available RNA-Seq reads from *S. hispanica* seeds (INSDC Accession PRJNA196477) (Sreedhar et al., 2015) and the leaf tissues (INSDC Accession PRJNA359989) (Boachon et al., 2018) were aligned against our assembled Chia transcriptome: about 69% sequence reads from the seed, and 43% from the leaf tissues were mapped to our Chia transcript assemblies.

Peptide sequences from the assembled transcripts were generated by using TransDecoder. About 65,587 of total 82,663 transcripts from Chia were translated into 99,307 peptides. The number of peptides is higher than the number of transcripts assembled due to the occurrence of multiple open reading frames (ORFs) in a single transcript. Functional annotation of peptides was first carried out by using InterProScan (Jones et al., 2014) to assign structural-functional domains and then carried out by employing the agriGO ontology enrichment analysis (Tian et al., 2017). We were successful in assigning InterPro accessions to the 45,209 peptides (**Supplementary Material 5**), and gene ontology (GO) terms to a total of 32,638 peptides (**Supplementary Material 6**). A total of 20,857 peptides were annotated with GO biological process (BP); 8,677 peptides were annotated with GO cellular component (CC) terms, and 26,877 peptides were annotated to GO molecular function (MF) terms (**Supplementary Material 6**).

The Vegetative and Reproductive Plant Structures of Chia Show Distinct Gene Expression Patterns

The RNA-Seq by expectation-maximization (RSEM) package (Li and Dewey, 2011) was used to calculate Fragments Per Kilobase of transcript per Million mapped reads (FPKM) value for the final set of 82,663 assembled transcripts from Chia. After removing the transcripts with an extremely low expression, we considered 82,385 transcripts for further analysis.

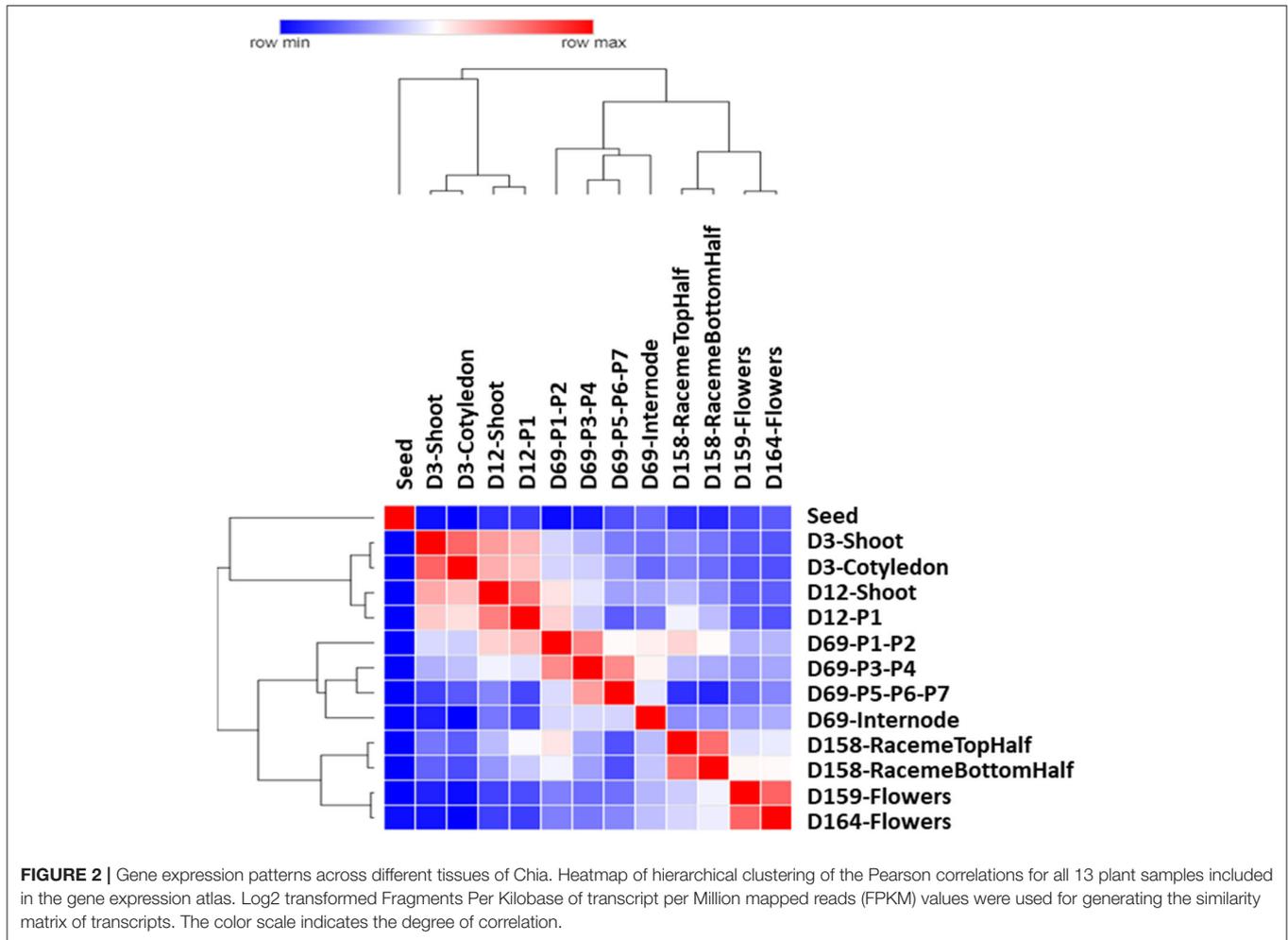
The hierarchical clustering of Pearson's correlations based upon the FPKM values of transcripts across all tissue samples



provides an insight into the spatial and temporal gene expression pattern (Figure 2). Based on their plant structure and developmental stage attributes, the samples, for example, vegetative tissues, D3 (cotyledon and shoot), and D12 (shoot and very first leaf P1 at shoot apex), tend to be clustered together. D69 leaf samples from the developmental stages P1–P7 tend to be clustered together along with the D69 internode, which was located between the P6 and P7 nodes on the main stem. Similarly, samples of the reproductive plant structure of flowers (D159 and D164) and inflorescence (raceme top and bottom half) tend to be clustered together.

Overall, 53,200 of 82,385 transcripts showed a differential expression across the 13 samples; however, only 38,480

transcripts show a significant difference (\log_2 fold change ≥ 2) (see Supplementary Material 7). A summary of differentially expressed transcripts (DETs) across all tissues is shown in Table 2. In general, all tissue types show higher numbers of downregulated DETs compared to upregulated DETs except D3-cotyledon. Moreover, 1,696 DETs are common in all 13 samples, and a majority of DETs are regulated in more than one sample. Closely related tissue types show a maximum overlap in their transcriptome. Figure 3 shows common and unique DETs among the plant structure types that are grouped based on developmental stages. The samples collected from the initial growth stages (D3-cotyledon, D3-shoot, and D12-shoot) share only 213 DETs with the seeds, which have a remarkably distinct



transcriptome (**Figure 3A**). Among all leaf developmental stages, only 368 DETs were common. Mature leaves (D69-P5-P6-P7) contain maximum DETs in both upregulated and downregulated categories (**Figure 3B**). Notably, in early leaf developmental stages (D12-P1 and D69-P1-P2), transcripts encoding for *Growth Regulating Factors* (*GRF2*, *GRF5*) and bHLH domain-containing (*SPEECHLESS*) transcription factors (TFs) were highly expressed. *GRF* TFs play an important role in leaf growth, and the bHLH *SPEECHLESS* factors are involved in stomata initiation and development (Kim et al., 2003; MacAlister et al., 2007; Kanaoka et al., 2008; Lampard et al., 2008). Among raceme inflorescence and flower tissues, 414 DETs were common and D164-Flowers had the maximum DETs (**Figure 3C**).

Based on the expression profile, Chia transcripts can be grouped into 20 co-expression clusters (**Supplementary Material 8**). Notably, cluster 1 is enriched in DETs that show preferential expression in the seeds; cluster 5, 12, and 20 are enriched in DETs that show preferential expression in the vegetative tissues; and cluster 15 and 18 contain DETs showing preferential expression in reproductive tissues (**Supplementary Material 9**).

The 7,507 transcripts in cluster 1 showed preferential expression in the seeds, including the transcripts coding for the late embryogenesis abundant (LEA) proteins, seed storage proteins, oil body-associated proteins, and oleosin family proteins (**Supplementary Material 7**). In contrast, cluster 5 (4,616 transcripts) is enriched with the transcripts that show preferential expression in D3-cotyledon, D3-shoot including those coding for the TFs of basic leucine zipper family, photosystem I and II proteins, aquaporins, and calcineurin-like phosphoesterase domain-containing proteins (**Supplementary Material 9**).

Cluster 12 includes the 3,909 transcripts that show preferential expression in the D69-P5-P6-P7 leaf stages, including the transcripts coding for leucine-rich receptor-like kinases (LRR-RLKs) and wall-associated receptor kinases (WAKs), root hair-defective 3 GTP-binding (*RHD3*) domain-containing proteins, such as ABC, phosphate, and aluminum transporter proteins, cytochrome P450s, glycosyltransferases, and *WRKY* TFs (**Supplementary Material 7**). Both LRR-RLKs and WAKs are known to play roles in disease resistance and abiotic stress response (Harkenrider et al., 2016; Al-Bader

et al., 2019; Wang et al., 2019; Amsbury, 2020; Zhang et al., 2020).

Cluster 20 consists of 2,173 transcripts that show preferential expression in the D69 Internode (**Supplementary Material 9**).

TABLE 2 | Differentially expressed transcripts (DETs) across various developmental stages.

Tissue type	Total DET	Tissue-specific	Upregulated (\log_2 FC ≥ 2)	Downregulated (\log_2 FC ≤ 2)
Seed	28,641	13,450	6,284	13,429
D3-Cotyledon	7,377	1,781	2,632	1,746
D3-Shoot	3,415	495	970	1,161
D12-Shoot	2,136	270	52	1,521
D12-P1	8,795	2,038	770	3,976
D69-P1-P2	3,511	633	288	2,185
D69-P3-P4	3,019	556	479	1,701
D69-P5-P6-P7	14,140	3,504	1,884	6,637
D69-Internode	9,260	2,152	1,390	5,163
D158-RacemeTopHalf	5,591	1,183	770	2,865
D158-RacemeBottomHalf	3,614	804	852	1,860
D159-Flowers	6,047	1,136	1,274	2,883
D164-Flowers	6,134	879	969	3,353

It includes transcripts encoding MYB (*MYB54*, *MYB52*) and NAC TFs, xyloglucan endotransglucosylase, and receptor-like kinases (RLKs). *MYB54*, *MYB52*, and NAC TFs are known for their role in the internode development and secondary cell wall biosynthesis (Zhong et al., 2008; Grant et al., 2010; Cassan-Wang et al., 2013). Xyloglucan endotransglucosylase and RLKs are also involved in cell wall biosynthesis and expansion (Guo et al., 2009; Haruta et al., 2014).

Cluster 15 consists of 3,619 transcripts that showed a high expression in flowers. It includes the transcripts coding for beta-glucosidase, multidrug and toxic compound extrusion transporter proteins, cinnamyl alcohol dehydrogenase, germin-like proteins, pectin acylesterases, *MYB21* and *MYB24* TFs, *ZFP2* TF, GDSL lipases, and cytochrome P450s (**Supplementary Material 9**). *MYB21* and *MYB24* TFs play a role in petal, stamen, and gynoecium development in flowers (Reeves et al., 2012), and *ZFP2* controls floral organ abscission (Cai and Lashbrook, 2008). Cinnamyl alcohol dehydrogenases are involved in lignin biosynthesis in floral stem in *Arabidopsis* (Sibout et al., 2005), and germin-like proteins play an important role in response to pathogens (Zimmermann et al., 2006; Manosalva et al., 2009; Wang et al., 2013). Also, the 2,679 transcripts in cluster 18 showed an upregulation in the inflorescence tissues. Agamous-like MADS-box proteins and MYB family TFs that play a vital role in floral meristem development are enriched in cluster

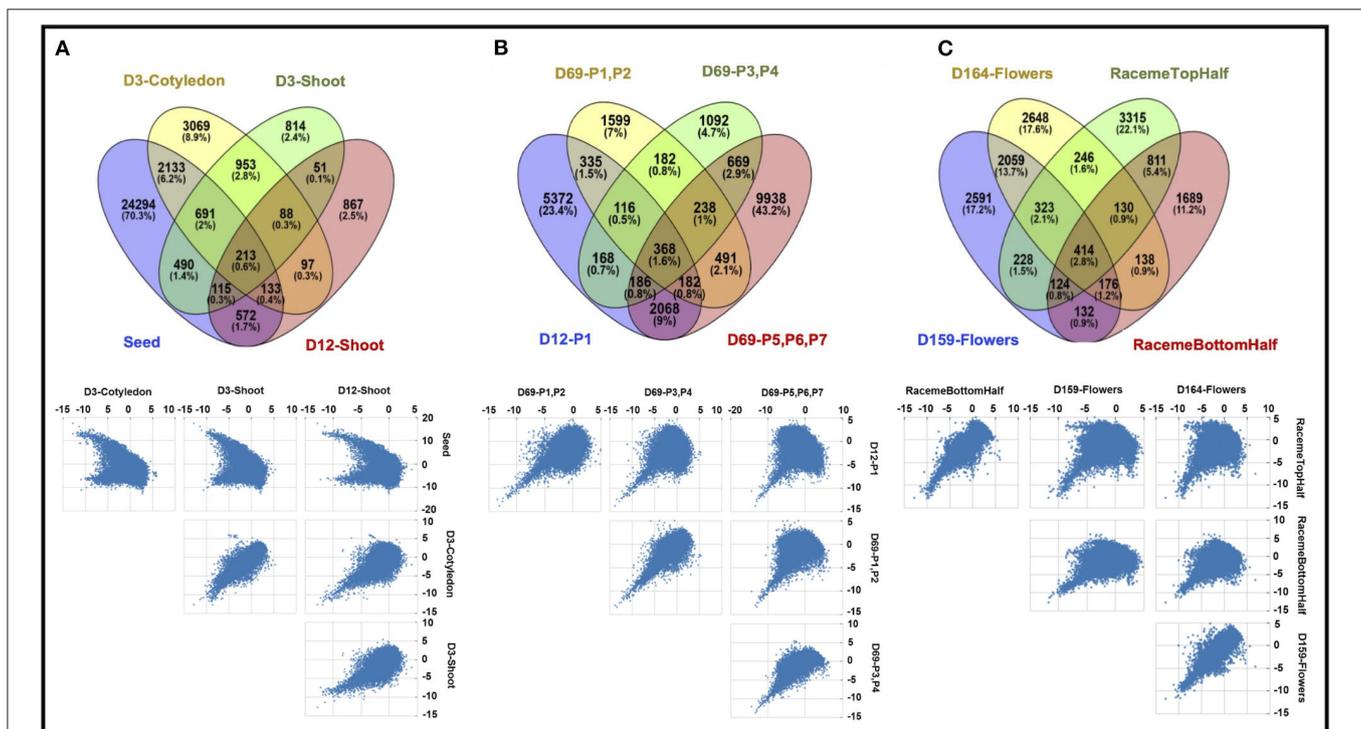
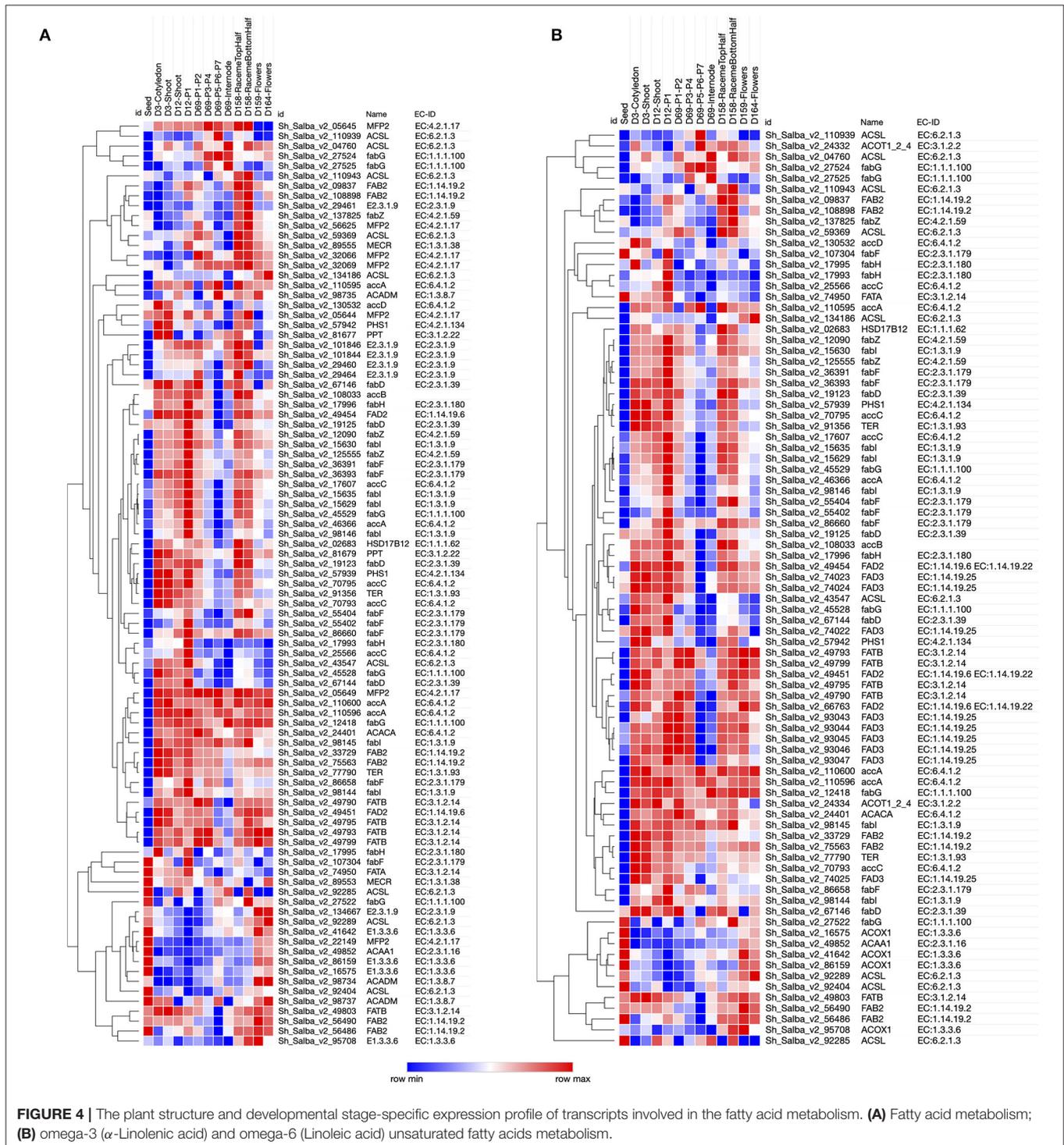


FIGURE 3 | Comparison of differentially expressed transcripts (DETs) of Chia among (A) seed, D3-cotyledon, D3-shoot, and D12-shoot; (B) leaf samples from D12-P1 and D69-P1-P2, D69-P3-P4, D69-P5-P6-P7; (C) reproductive stage samples, including D158 inflorescence raceme top and bottom half, D159- and D164-flowers. The Venn diagrams in the upper panel represent common and unique DETs in each sample, and scatter plots in the lower panel represent the distribution pattern of DETs.

18 (Supplementary Material 9). MYBs and MADS-box TFs are the essential regulators of various developmental processes (Zimmermann et al., 2004; Millar and Gubler, 2005; Yang et al., 2007; Gomez et al., 2011; Kobayashi et al., 2015). We find that transcripts annotated as terpene synthases show an upregulation in flowers as compared to the inflorescence tissues.

Biological Pathways Enriched Across Different Development Stages

The metabolic network representation across the developmental stages of Chia was determined by mapping to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. We were successful in mapping a total of 5,555 Chia transcripts to



464 KEGG pathways, including starch and sucrose metabolism (PATH:ko00500), fatty acid metabolism (PATH:ko01040), phenylpropanoid biosynthesis (PATH:ko00940), photosynthesis (PATH:ko00195), fatty acid biosynthesis (PATH:ko00061), and amino acids metabolism (**Supplementary Material 10**). The expression pattern of transcripts encoding the enzymes for fatty acid metabolism and PUFA metabolism across different developmental stages was analyzed (**Figure 4**). Transcripts encoding acetyl-CoA carboxylase (EC 6.4.1.2), the very first enzyme catalyzing the conversion of acetyl-CoA to malonyl-CoA in the fatty acid biosynthesis, were highly expressed in all tissues except seeds. In the next reaction of this pathway, the malonyl group from malonyl-CoA is transferred to acyl carrier proteins (ACPs) for further elongation. We identified the transcripts for all the enzymes participating in the elongation steps. Acyl-ACP thioesterases (3.1.2.14) act in the last steps of fatty acid biosynthesis and serve as a determining factor for the generation of a variety of fatty acids within an organism. We further analyzed the expression of transcripts encoding for enzymes associated with PUFA metabolism (**Figure 4B**). We identified 32 fatty acid desaturase (FAD) transcripts from *FAD2*, *FAD3*, *FAD6*, *FAD7*, and *FAD8* families (**Table 3**). FADs are critical for catalyzing the fatty acid desaturation. Endoplasmic reticulum-localized *FAD2* and plastid-localized *FAD6* encode two ω -6 desaturases required to convert oleic acid to linoleic acid (18:2 $\Delta^{9,12}$) (Zhang et al., 2012). The desaturation of linoleic acid (18:2 $\Delta^{9,12}$) to α -linolenic acid (18:3 $\Delta^{9,12,15}$) is catalyzed by the endoplasmic reticulum-localized *FAD3* and plastid-localized *FAD7* and *FAD8* proteins (Dar et al., 2017; Xue et al., 2018).

Mint, a Lamiaceae family plant, is primarily known for the production of monoterpenes, e.g., menthol and limonene (Aharoni et al., 2005; Ahkami et al., 2015); however, the majority of Chia terpenes are sesqui-, di-, and tri-terpenes (Ma et al., 2012; Cui et al., 2015; Triikka et al., 2015). In our Chia dataset, we observed the expression profile of transcripts involved in the biosynthesis of terpenoid backbone, monoterpenes, and sesquiterpenes. Transcripts encoding the enzymes of the 2-C-methyl-D-erythritol 4-phosphate (MEP) and the mevalonate (MVA) pathways involved in the terpenoid backbone biosynthesis showed differential expression patterns among all tissue types (**Figure 5**). Transcripts for monoterpene synthases, such as 1,8-cineole synthase (EC 4.2.3.108), myrcene synthase (EC 4.2.3.15), and linalool synthase (EC 4.2.3.25), were highly expressed in the reproductive stage samples (**Figure 5**), indicating that flowers are the prime site for the biosynthesis of essential oils known to have therapeutic properties. However, transcripts for the sesquiterpene synthases, β -caryophyllene synthase (EC 4.2.3.57), α -humulene synthase (EC 4.2.3.104), germacrene synthase (EC 4.2.3.60), and solavetivone oxygenase (EC 4.2.3.21) were enriched in the vegetative tissues (**Figure 5**). Sesquiterpene, β -caryophyllene, α -humulene, Germacrene, and solavetivone are likely to play a role in herbivory defense.

Identification of Distinct Tissue-Specific Networks of TFs in Chia

Transcription factors are the key regulators of a plant's growth and development. We identified 633 DETs encoding TFs

TABLE 3 | Transcripts annotated as Chia fatty acid desaturase (FAD).

Fatty acid desaturases	Transcripts
<i>FAD2</i>	Sh_Salba_v2_49454
	Sh_Salba_v2_49451
	Sh_Salba_v2_66763
<i>FAD3</i>	Sh_Salba_v2_74023
	Sh_Salba_v2_93044
	Sh_Salba_v2_93043
	Sh_Salba_v2_74025
	Sh_Salba_v2_93046
	Sh_Salba_v2_74024
	Sh_Salba_v2_74022
	Sh_Salba_v2_93047
	Sh_Salba_v2_93045
	<i>FAD6</i>
Sh_Salba_v2_05731	
Sh_Salba_v2_05730	
Sh_Salba_v2_05725	
Sh_Salba_v2_05728	
Sh_Salba_v2_05721	
Sh_Salba_v2_05724	
<i>FAD7</i>	Sh_Salba_v2_69172
<i>FAD8</i>	Sh_Salba_v2_69173
	Sh_Salba_v2_90850
	Sh_Salba_v2_52578
	Sh_Salba_v2_69162
	Sh_Salba_v2_52570
	Sh_Salba_v2_52575
	Sh_Salba_v2_52576
	Sh_Salba_v2_69169
	Sh_Salba_v2_69166
	Sh_Salba_v2_52573
Sh_Salba_v2_69174	
Sh_Salba_v2_69171	

belonging to 53 families (**Supplementary Material 11**) in our Chia transcriptome. The highest number of transcripts belongs to *MYB* (60), followed by *bHLH* (45), *NAC* (38), *bZIP* (32), *WRKY* (28), *C2H2* (27), *MYB-related* (25), *MADS-box* (26), *C3H* (24), *G2-like* (22), *Hd-ZIP* (22), *Trihelix* (17), *TCP* (14), *Dof* (13), *GATA* (13), *GRAS* (13), and *TALE* (13) gene families. The expression pattern of differentially expressed TFs across the developmental stages is shown in **Supplementary Material 12A**. To build a co-expression network, we first filtered out highly upregulated TFs in any of the 13 samples ($\geq 5 \log_2$ fold change) and then used 23 TFs as baits (nodes) and the FPKM value of 38,480 transcripts as an expression matrix. The seed-specific TF network showed 95,318 connections among 21 TFs and 4,629 transcript nodes (**Supplementary Materials 12A, 13**). A set of 92 transcripts (green) are connected to all 21 TFs. Two sets of 15 transcripts (teal) and 33 transcripts (pink) are connected to an ethylene responsive factor (ERF) (T112851) that is highly upregulated (\log_2 fold change 5.561) in the seeds. Most of the transcripts correlated to ERF (T112851) are downregulated

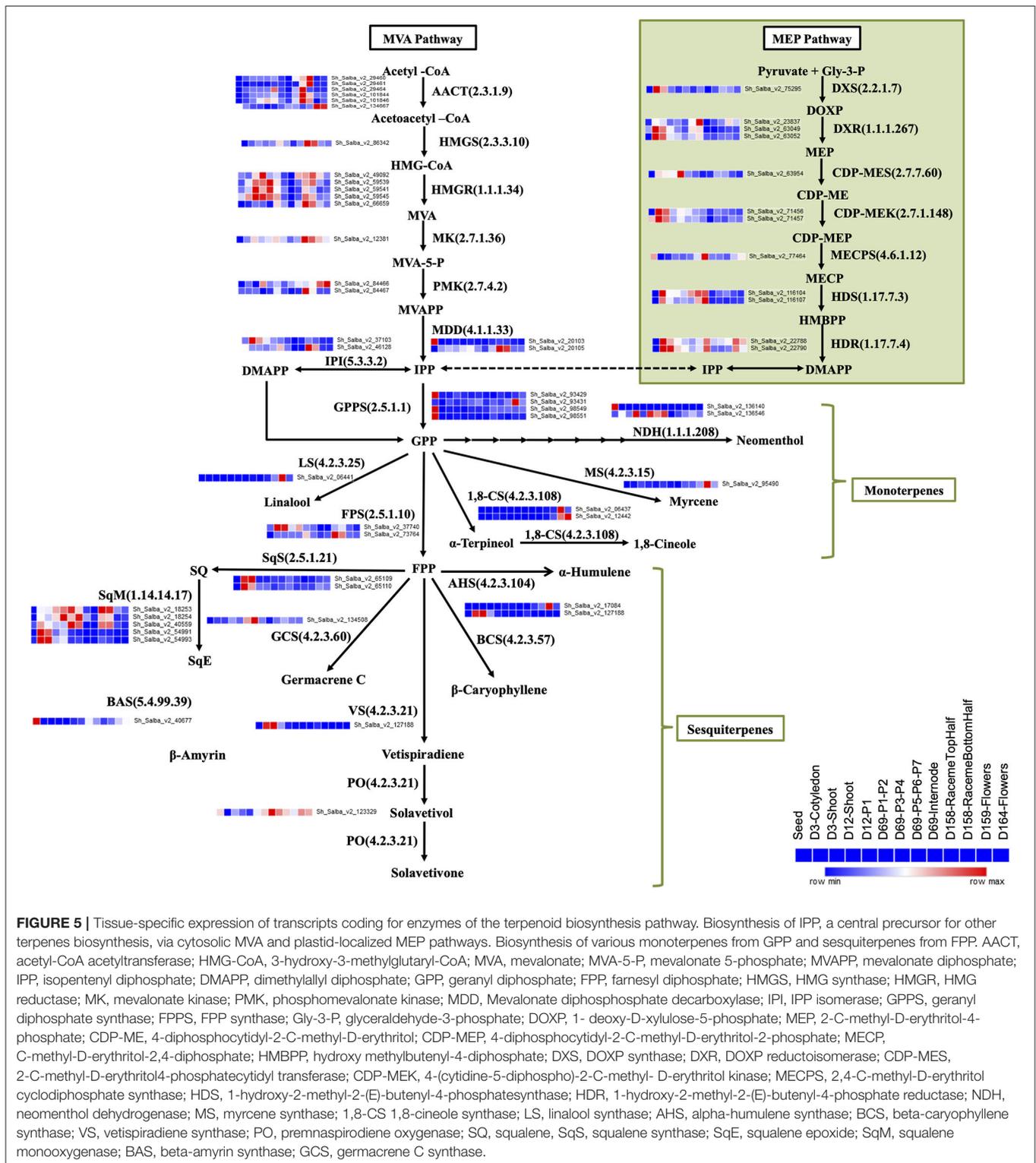


FIGURE 5 | Tissue-specific expression of transcripts coding for enzymes of the terpenoid biosynthesis pathway. Biosynthesis of IPP, a central precursor for other terpenes biosynthesis, via cytosolic MVA and plastid-localized MEP pathways. Biosynthesis of various monoterpenes and sesquiterpenes from GPP. AACT, acetyl-CoA acetyltransferase; HMGS-CoA, 3-hydroxy-3-methylglutaryl-CoA; MVA, mevalonate; MVA-5-P, mevalonate 5-phosphate; MVAPP, mevalonate diphosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GPP, geranyl diphosphate; FPP, farnesyl diphosphate; HMGS, HMG synthase; HMGR, HMG reductase; MK, mevalonate kinase; PMK, phosphomevalonate kinase; MDD, Mevalonate diphosphosphate decarboxylase; IPI, IPP isomerase; GPPS, geranyl diphosphate synthase; FPS, FPP synthase; Gly-3-P, glyceraldehyde-3-phosphate; DXS, 1-deoxy-D-xylulose-5-phosphate; MEP, 2-C-methyl-D-erythritol-4-phosphate; CDP-ME, 4-diphosphocytidyl-2-C-methyl-D-erythritol; CDP-MEP, 4-diphosphocytidyl-2-C-methyl-D-erythritol-2-phosphate; MECP, C-methyl-D-erythritol-2,4-diphosphate; HMBPP, hydroxy methylbutenyl-4-diphosphate; DXS, DOXP synthase; DXR, DOXP reductoisomerase; CDP-MES, 2-C-methyl-D-erythritol-4-phosphatecytidyl transferase; CDP-MEK, 4-(cytidine-5-diphospho)-2-C-methyl-D-erythritol kinase; MECPS, 2,4-C-methyl-D-erythritol cyclodiphosphate synthase; HDS, 1-hydroxy-2-methyl-2-(E)-butenyl-4-phosphatesynthase; HDR, 1-hydroxy-2-methyl-2-(E)-butenyl-4-phosphate reductase; NDH, neomenthol dehydrogenase; MS, myrcene synthase; 1,8-CS, 1,8-cineole synthase; LS, linalool synthase; AHS, alpha-humulene synthase; BCS, beta-caryophyllene synthase; VS, vetispiradiene synthase; PO, premnaspirodiene oxygenase; SQ, squalene, SqS, squalene synthase; SqE, squalene epoxide; SqM, squalene monooxygenase; BAS, beta-amyryn synthase; GCS, germacrene C synthase.

in the seed and other tissue types. Two transcripts (T135081 and T106468) exclusively connected to NIN-like TF (T145088) are downregulated in the seeds and other tissues. A bZIP TF showed a connection to a transcript T120275, which is

upregulated in the seeds but downregulated in other tissues. The MYB TF (T131530) is upregulated in the seeds and connected to a set of 34 transcripts (magenta) that were mostly downregulated in the seeds and other tissues. Bait transcripts

are also correlated to each other, suggesting multiple regulatory modules within the network (**Supplementary Material 12**). Two TFs, *MYB* (T130985/Sh_Salba_v2_130985) and *C3H* (T121906/Sh_Salba_v2_121906), which were highly expressed in the D69-Internode but downregulated or absent in other tissues, showed connections (edges) to a set of 504 transcripts with no connection to any other bait nodes (**Supplementary Material 12B**).

Identification of SSRs Markers

Simple sequence repeats are an important class of genetic markers widely used in molecular breeding applications. The identified SSRs from the transcriptome are highly advantageous as compared to the identified SSRs from the genome sequence. If an SSR from the transcribed region is polymorphic, it may have a direct impact on the expression, structure, stability of a transcript, and/or on a peptide sequence and the functional domains. We identified a total of 2,411 SSRs in the *de novo* assembled transcriptome represented by di-, tri-, and tetra-nucleotide motifs (**Supplementary Material 14**). The most abundant di-, tri-, and tetra-nucleotide motifs were CT (201), GAA (84), and AGTC (12), respectively (**Supplementary Material 15**). A total of 1,771 SSRs were present in the DETs, and 148 SSRs were found in the expressed transcripts mapped to at least one metabolic pathway (**Supplementary Material 16**).

DISCUSSION

To develop a repertoire of genetic and genomic data and knowledge resources on Chia, recent limited efforts have focused on the expression of lipid biosynthesis and terpene synthase genes in developing Chia seeds, roots, and leaves (Sreedhar et al., 2015; Peláez et al., 2019; Wimberley et al., 2020). We developed a comprehensive gene expression atlas for Chia from the 13 different plant structure types collected at various developmental stages using RNA-Seq coupled with the *de novo* transcriptome assembly approach (**Table 1**). This transcriptome atlas covers the complete assembly of about 86% of the genes present in the Chia genome. The assembled transcripts were annotated by using BLASTx and tBLASTx and translated into peptides using Transdecoder (v2.1.0); the peptides were annotated with GO terms and structurally conserved domains. Overall, the Chia transcriptome data set is comprehensive and covers the majority of genes participating in the cellular metabolic process, catalytic activity, regulation of gene expression, transport, ion binding, organelle development and function, and formation of macromolecular complexes. Thus, we present a much needed reference resource for the breeding and improvement of this important crop.

In the comparison of Chia transcripts datasets to genomic/transcriptome data sets from the six most closely related eudicots (**Figure 1C**), we found the topmost matching of a Chia transcript with the transcripts of perennial herbs, the red sage *S. miltiorrhiza* (Wenping et al., 2011) and the scarlet sage, *S. splendens* (Ge et al., 2014)—both species are

rich in secondary metabolites and are used in traditional medicine. Furthermore, the Chia expression atlas provides insights into the functional relatedness of genes and their expression across the developmental stages and tissue types. Hierarchical clustering of Chia transcripts suggests the role of different gene families in the development of each growth stage, thus providing a foundation for studying the molecular mechanisms and metabolic pathways occurring in different tissues and developmental stages. For example, GRF family TFs, likely to play an essential role in regulating leaf number and size show high expression in D69-P1-P2 leaf stages. GRF family TFs play an essential role in the growth and development of leaf, and likewise express highly in D69-P1-P2 leaf stages. In *Arabidopsis*, *GRF1*, *GRF2*, and *GRF5* regulate leaf number and size (Kim et al., 2003; Horiguchi et al., 2005; Lee et al., 2009). Likewise, in mature leaves, transcripts coding for LRR-RLKs and WAKs proteins show high expression, which is consistent with their role in guard cells and stomatal patterning (Shpak et al., 2005) and biotic resistance (Harkenrider et al., 2016).

Essential oils, the secondary metabolites of the terpenoid biosynthesis pathways, are highly desired for their usage in medicine, food, and cosmetics, and have a potential survival benefit for the plant against insects, herbivores, and pathogens. In this Chia dataset, we identified transcripts encoding enzymes for the terpenoid backbone biosynthesis (**Figure 5**). Monoterpene synthases are involved in the biosynthesis of essential oils, and sesquiterpene synthases primarily contribute to the biosynthesis of insecticidal compounds. Chia flowers show a higher expression of monoterpene synthase transcripts (**Figure 5**) that code for 1, 8-cineole synthase (EC 4.2.3.108) and β -myrcene synthase (EC 4.2.3.15). Cineole and myrcene are found in fragrant plants and are known to have therapeutic medicinal properties such as sedative, anti-inflammatory, antispasmodic, and antioxidant (do Vale et al., 2002; Moss and Oliver, 2012; Bouajaj et al., 2013; Juergens, 2014; Khedher et al., 2017). The comparison of reproductive vs. vegetative tissue shows that monoterpene synthases were expressed highly in the reproductive tissues, and sesquiterpene synthases were prominent in the vegetative tissues. These findings confirm that flowers are involved in the synthesis of fragrant and essential oils whereas the vegetative tissues are rich in compounds, which are known for their herbivory defense and insecticidal properties. Phenylpropanoid and flavonoid biosynthesis pathways are also highly enriched in seeds and other tissue types (**Supplementary Material 10**).

Chia seeds are a rich source of PUFAs. We observed a lower expression of FAD transcripts in the seeds in comparison to other tissue types (**Figure 4**), suggesting that the seeds might serve as a storage organ for PUFAs rather than the synthesis site.

The correlation analysis hinted a significant relationship between highly upregulated TFs and the other DETs (**Supplementary Material 12**). The co-expression analysis suggested 21 TFs that are members of the B3, bZIP, ERF, WOX, AP2, MYB, C3H, EIL, LBD, DBB, NIN-like, and HSF families, play critical roles in the regulation of target gene expression across various developmental stages. We also observed that the two MYB and C3H zinc finger TFs were highly upregulated

in the D69-Internode (**Supplementary Material 12**). Their expression is consistent with the reports on their orthologs' role in internode elongation and development processes (Zhong et al., 2008; Kebrom et al., 2017; Gómez-Ariza et al., 2019). Of the 21 TFs highly expressed in the seed samples, T112851 is an AP2/ERF family member. Its homologs are known to play a role in dehydration-induced response as the DREB2A proteins that are involved in response to drought, salt, and low-temperature stress (Nakashima et al., 2000; Sakuma et al., 2002). We expected to see such stress-responsive genes in seed samples because the seeds undergoes dehydration during their maturation (Naithani et al., 2017).

The conventional use of the BUSCO gene analysis is to assess the completeness of genome and transcriptome. However, in the first report of its kind known to us, we used the Chia BUSCO gene set to assess the gene expression data quality. A baseline transcript abundance or the expression levels of a conserved Viridiplantae BUSCO gene set were compared between Chia and *Arabidopsis* (**Supplementary Materials 17, 18**). In the Chia transcriptome atlas data set, 601 transcripts were mapped to 411 BUSCO genes. They showed sample-specific transcript abundance: 49 Chia transcripts and their *Arabidopsis* homologs show a similar higher abundance in dry seeds, and 28 transcripts in the D69 leaves and their *Arabidopsis* homologs show similar transcript abundance profiles in leaf and plant parts carrying leaf-like structures. The conserved expression profiles of homologous BUSCO genes in taxonomically diverse Chia and *Arabidopsis* plants (**Supplementary Material 17**) support the high-quality of Chia transcriptome generated in this study. Our gene expression validation analysis carries more genes (>400), uses the conserved green plant BUSCO gene set, and improves the typical use of the reverse transcription (RT)-PCR method to quantify 5–10 genes for validating transcript abundance. This analysis was only possible due to the availability of high-quality publicly available transcriptome data sets provided by the EMBL-EBI Gene Expression Atlas (<https://www.ebi.ac.uk/gxa/home>) (Papatheodorou et al., 2018), supported by the Ensembl Plants (<http://plants.ensembl.org/index.html>) (Howe et al., 2020) and the Gramene databases (<http://www.gramene.org>) (Tello-Ruiz et al., 2021).

Further analysis of the *de novo* assembled Chia transcriptome revealed 2,411 SSRs (**Supplementary Material 15**). SSRs are an important class of genetic markers widely used in molecular breeding applications. The identified SSRs from the transcriptome are highly advantageous as compared to the identified SSRs from the genome because they are from the transcribed region, and if they are polymorphic, they may have a direct impact on the expression, structure, and stability, of the transcript and the peptide (**Supplementary Material 16**). The SSRs identified in the Chia reference transcriptome are a valuable resource for breeding and genetic improvement of the crop.

Overall, this is the first study on the generation of a well-annotated plant structure-specific reference transcriptome atlas for Chia, a neo model, and an agronomically important crop. We expect that the raw and analyzed Chia transcriptome sequence data and 2,411 SSRs identified in this study would serve as

an important resource for the researchers working on Chia and other important plant species of the mint family. The transcriptome data will greatly help in correcting errors in the future genome assembly of *S. hispanica*, the identification of gene models, improving the gene and genome annotation, and the development of a Chia-specific metabolic network. This transcriptome study is expected to initiate opportunities to undertake comparative and functional genomics, pathway analyses, and genome to phenome studies in Chia.

MATERIALS AND METHODS

Plant Material, Growth Conditions, and Sampling

Seeds of Chia (*S. hispanica* L.) were bought online from Ancient Naturals, LLC, Salba Corp., N.A., were sown in autoclaved soils, and watered thoroughly under controlled greenhouse conditions. All seeds germinated on the third day after sowing (DAS). Since the primary seed material was expected to be a heterogeneous mixture, biological replicates for each tissue type were collected from three randomly chosen plants. The description of the samples collected from various developmental stages and tissue types is shown in **Table 1**. The tissue samples include seeds, green cotyledons, shoots after 3 and 12 DAS, leaves from day-12 (D12-P1) and day-69 DAS (D69-P1-P7), an internode (D69-Internode) between the P5 and P6 leaf nodes collected on day-69 DAS, top and bottom halves of the raceme inflorescence from 158 DAS (D158) carrying pre-anthesis flowers, and flowers from the day-1 and day-5 of flowering (anthesis stage) (D159) and 164 (D164) DAS. Collected samples were immediately frozen in liquid nitrogen and stored at -80°C .

Sample Preparation and Sequencing

Total RNA from the frozen tissues was extracted as per the manufacturer's protocol using the RNA Plant reagent (Invitrogen Inc., Waltham, MA, USA), RNeasy kits (Qiagen Inc., Germantown MD, USA), and treatment with RNase-free DNase (Life Technologies Inc., Carlsbad, CA, USA). Total RNA concentration and quality were determined by using ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and Bioanalyzer 2100 (Agilent Technologies Inc., Santa Clara, CA, USA). Samples were prepared separately from each of the three biological replicates of each tissue type by using the TruSeq™ RNA Sample Preparation Kits (v2) and sequenced by using the Illumina HiSeq 2500 instrument (Illumina Inc., San Diego, CA, USA) at the Center for Genomic Research and Biocomputing, Oregon State University, Corvallis, OR, USA.

De novo Transcriptome Assembly and Quality Estimation

FASTQ file generation from the RNA-Seq sequences was done by using the CASAVA software v1.8.2 (Illumina Inc., San Diego, CA, USA). Read quality was assessed by using FastQC, and poor-quality reads were removed with Sickle v. 1.33 ($-q = 20$) ("najoshi/sickle"). The transcripts were assembled by using Velvet (v1.2.10), which uses De Bruijn graphs to assemble short reads (Zerbino and Birney, 2008). An assembly of 67 and 71

k-mer lengths was performed by using all tissue-specific reads. The assemblies produced by Velvet were merged into a single consensus assembly by Oases (v0.2.08) (Schulz et al., 2012), which produced transcript isoforms using read sequence and pairing information. Quality estimation for reducing redundancy in a transcript assembly (a quality control check for *de novo* assembled transcriptome) was carried out by using CD-HIT-EST (Li and Godzik, 2006), TransRate (Smith-Unna et al., 2016), and QUAST (Gurevich et al., 2013) software packages. The assembled transcripts passing the CD-HIT-EST quality control step were used for further downstream analyses and considered a reference transcriptome for differential gene expression analyses. The BUSCO analysis (Simão et al., 2015) on the Chia transcriptome was also performed at the Galaxy platform (Afgan et al., 2018) to assess the completeness of the transcriptome set and coverage of the Chia gene set. We used the BUSCO version 4.1.2 and the Viridiplantae (green plant) lineage-specific data set viridiplantae_odb10, with a 425 reference single-copy core gene set.

Functional Annotation and Pathway Enrichment Analysis

Assembled transcripts were annotated by using BLASTx and tBLASTx with a value of *E* cutoff of 10^{-10} . The assembled transcripts were translated into peptides by using Transdecoder (v2.1.0) [“TransDecoder (Find Coding Regions Within Transcripts)”] with a minimum peptide length of 50 or more amino acids. Transdecoder used the BLASTp and PfamA search results to predict the translated ORF. The resulting peptides were analyzed by using the InterProScan Sequence Search (v5.17.56) (Jones et al., 2014) hosted by the Discovery Environment and powered by CyVerse (Joyce et al., 2017). We used the agriGO Analysis Toolkit (Tian et al., 2017) to identify statistically enriched function groups of transcripts. agriGO uses a Fisher’s exact test with a Yekutieli correction for the false discovery rate calculation. Significance cutoffs were set at a value of *p* equal to 0.05 and a minimum of five mapping entries per GO term. KAAS-KEGG automation server was used for an ortholog assignment and a pathway analysis (Moriya et al., 2007).

Gene Expression and Clustering

Bowtie2 (Langmead and Salzberg, 2012) was used to align the sequence reads from each tissue type to the assembled transcriptome. The RSEM software package (Li and Dewey, 2011) was used to estimate the transcript expression counts (FPKM) from the aligned sequence reads. The count data obtained from RSEM was used in EBSeq (Leng et al., 2013) to identify differentially expressed genes based on the false discovery rate corrected value of *p* is 0.05. Heatmaps were generated by using Morpheus (<https://software.broadinstitute.org/morpheus>) and MEV (version 4.8.1) (MEV: MultiExperiment Viewer, 2017) to cluster the expression data from Chia. Log₂ transformed fold change value for each transcript was used as an input (value of *p* being 0.1). Due to the orders of magnitude in the expression of transcripts between the tissue types, we chose several data normalization methods for cluster generation. Unit variance, median centering of transcripts, and the summation of squares

were applied to the dataset. In the investigation of individual gene families, transcripts were hierarchically clustered using a Pearson correlation. The Chia transcripts were mapped to the BUSCO genes, and their *Arabidopsis* homologs were compared by using the *Arabidopsis* dataset E-MTAB-7978 (Mergner et al., 2020) from the EMBL-EBI Gene Expression Atlas (<https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-7978>).

Co-expression and Network Analysis

The TF transcripts were classified based on homology searches in Plant TFDB database v5.0 (<http://planttfdb.cbi.pku.edu.cn>) (Jin et al., 2017) and BlastX searches against *A. thaliana*. For the co-expression analysis, the CoExpNetViz tool (Tzfadia et al., 2015) was used. This tool utilizes a set of query or bait genes as an input and a gene expression dataset. TF transcripts displaying a maximum expression cutoff of log₂ transformed FPKM ≥ 5 were used as baits, and DETs displaying a maximum expression cutoff of log₂ transformed FPKM ≥ 2 were used as an expression matrix. Baits and an expression matrix were loaded in the CoExpNetViz tool, and the analysis was run on default parameters to calculate co-expression with the setting of the Pearson correlation coefficient. For an expression matrix, the transcripts are considered as co-expressed if their correlation does not lie between the lower (1st) and upper (99th) percentile of the distribution of correlations between a sample of genes per gene expression matrix. The output files from the CoExpNetViz tool were used for displaying a gene-co-expression network by using Cytoscape (version 3.7.1).

Identification of SSRs

Multiple-length nucleotide SSRs were identified in the transcripts of the CD-HIT-EST assembly by using the stand-alone version of the SSR Identification Tool (SSRIT) (Temnykh, 2001).

DATA AVAILABILITY STATEMENT

The raw sequencing data from all cDNA libraries were deposited at EMBL-EBI ArrayExpress under the experiment number E-MTAB-5515 (<https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5515>). All the analyzed data for this project is accessible from the Chia Genomics Database (ChiaGDB: <http://salvia.cgrb.oregonstate.edu/>) and the Pankaj Jaiswal (2017). Comparative analysis of reference transcriptome atlas and insight into essential fatty acids and terpenoid biosynthesis pathways from Chia (*Salvia hispanica*). CyVerse Data Commons. DOI 10.7946/P2192W.

AUTHOR CONTRIBUTIONS

PJ and MG conceptualized the project. EH and JP helped MG in determining growth conditions, maintaining plants, sample collection, and extracting RNA. PG, MG, and SN did the data analysis. PG, MG, SN, and PJ wrote the manuscript. All authors reviewed and approved the manuscript.

FUNDING

This work was supported by the funds provided to PJ by the Department of Botany and Plant Pathology in the College of Agricultural Sciences at Oregon State University. PG, MG, SN, and PJ were also supported by the National Science Foundation awards 1127112 and 1340112.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.667678/full#supplementary-material>

Supplementary Material 1 | A summary of the raw and clean reads obtained after the sequencing and preprocessing, respectively, and reads aligned to the reference transcriptome.

Supplementary Material 2 | Quality assessment of merged (column 2), CD-HIT-EST (column 3), and TransRate (column 4) assemblies using QUAST.

Supplementary Material 3 | Workflow of Chia transcriptome sequencing and downstream analysis.

Supplementary Material 4 | Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis of assembled Chia transcriptome.

Supplementary Material 5 | Functional annotation of Chia peptides using InterProScan.

Supplementary Material 6 | Gene ontology (GO) annotations of Chia peptides.

Supplementary Material 7 | Tissue-specific transcripts and their differential expression data for all Chia transcripts.

Supplementary Material 8 | *k*-means clustering of transcripts depicting tissue-specific gene expression across different developmental stages. The *y*-axis in each cluster denotes the mean-centered log₂ transformed FPKM values ranging from +17 to -17.

Supplementary Material 9 | Chia transcripts clustered in 20 clusters.

Supplementary Material 10 | Transcripts mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

Supplementary Material 11 | Differentially expressed transcription factors (TFs) across various developmental stages.

Supplementary Material 12 | Seed- and internode-specific transcription network. **(A)** The seed-specific TF network shows connections among 21 TFs and 4,629 transcripts (targets) and **(B)** the internode-specific TFs network shows connections among 2 TFs and 504 transcript (targets).

Supplementary Material 13 | Transcription network in seed and internode.

Supplementary Material 14 | Frequency distribution of simple sequence repeats (SSRs) in Chia transcripts.

Supplementary Material 15 | SSR motifs in Chia transcripts.

Supplementary Material 16 | SSRs identified in transcripts coding for enzymes.

Supplementary Material 17 | Heat maps of transcript abundance (log₂ FPKM) for **(A)** all BUSCO transcripts in Chia, **(B)** a comparison of seed-enriched Chia BUSCO transcripts and their *Arabidopsis* homologs, and **(C)** a comparison of leaf-enriched Chia BUSCO transcripts and their *Arabidopsis* homologs.

Supplementary Material 18 | FPKM values of Chia BUSCO transcripts across different developmental stages and their *Arabidopsis* homologs.

REFERENCES

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible, and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544. doi: 10.1093/nar/gky379
- Aharoni, A., Jongsma, M. A., and Bouwmeester, H. J. (2005). Volatile science? metabolic engineering of terpenoids in plants. *Trends Plant Sci.* 10, 594–602. doi: 10.1016/j.tplants.2005.10.005
- Ahkami, A., Johnson, S. R., Srividya, N., and Lange, B. M. (2015). Multiple levels of regulation determine monoterpenoid essential oil compositional variation in the mint family. *Mol. Plant* 8, 188–191. doi: 10.1016/j.molp.2014.11.009
- Al-Bader, N., Meier, A., Geniza, M., Gongora, Y. S., Oard, J., and Jaiswal, P. (2019). Loss of premature stop codon in the Wall-Associated Kinase 91 (OsWAK91) gene confers sheath blight disease resistance in rice. *bioRxiv [Preprint]* 625509. doi: 10.1101/625509
- Amato, M., Caruso, M. C., Guzzo, F., Galgano, F., Commisso, M., Bochicchio, R., et al. (2015). Nutritional quality of seeds and leaf metabolites of Chia (*Salvia hispanica* L.) from Southern Italy. *Eur. Food Res. Technol.* 241, 615–625. doi: 10.1007/s00217-015-2488-9
- Amsbury, S. (2020). Sensing attack: the role of wall-associated kinases in plant pathogen responses. *Plant Physiol.* 183, 1420–1421. doi: 10.1104/pp.20.00821
- Ayerza (h), R., and Coates, W. (2009). Influence of environment on growing period and yield, protein, oil and α -linolenic content of three Chia (*Salvia hispanica* L.) selections. *Ind. Crops Prod.* 30, 321–324. doi: 10.1016/j.indcrop.2009.03.009
- Baginsky, C., Arenas, J., Escobar, H., Garrido, M., Valero, N., Tello, D., et al. (2016). Growth and yield of Chia (*Salvia hispanica* L.) in the Mediterranean and desert climates of Chile. *Chil. J. Agric. Res.* 76, 255–264. doi: 10.4067/S0718-58392016000300001
- Boachon, B., Buell, C. R., Crisovan, E., Dudareva, N., Garcia, N., Godden, G., et al. (2018). Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Mol. Plant* 11, 1084–1096. doi: 10.1016/j.molp.2018.06.002
- Bouajaj, S., Benyamna, A., Bouamama, H., Romane, A., Falconieri, D., Piras, A., et al. (2013). Antibacterial, allelopathic, and antioxidant activities of essential oil of *Salvia officinalis* L. growing wild in the Atlas Mountains of Morocco. *Nat. Prod. Res.* 27, 1673–1676. doi: 10.1080/14786419.2012.751600
- Cahill, J. P. (2005). Human selection and domestication of Chia (*Salvia hispanica* L.). *J. Ethnobiol.* 25, 155–174. doi: 10.2993/0278-0771(2005)25155:HSADOC2.0.CO;2
- Cai, S., and Lashbrook, C. C. (2008). Stamen abscission zone transcriptome profiling reveals new candidates for abscission control: enhanced retention of floral organs in transgenic plants overexpressing Arabidopsis ZINC FINGER PROTEIN2. *Plant Physiol.* 146, 1305–1321. doi: 10.1104/pp.107.110908
- Cañas, R. A., Li, Z., Pascual, M. B., Castro-Rodríguez, V., Ávila, C., Sterck, L., et al. (2017). The gene expression landscape of pine seedling tissues. *Plant J. Cell Mol. Biol.* 91, 1064–1087. doi: 10.1111/tpj.13617
- Cassan-Wang, H., Goué, N., Saidi, M. N., Legay, S., Sivadon, P., Goffner, D., et al. (2013). Identification of novel transcription factors regulating secondary cell wall formation in Arabidopsis. *Front. Plant Sci.* 4:189. doi: 10.3389/fpls.2013.00189
- Chicco, A. G., D'Alessandro, M. E., Hein, G. J., Oliva, M. E., and Lombardo, Y. B. (2009). Dietary Chia seed (*Salvia hispanica* L.) rich in α -linolenic acid improves adiposity and normalises hypertriglycerolaemia and insulin resistance in dyslipaemic rats. *Br. J. Nutr.* 101, 41–50. doi: 10.1017/S000711450899053X
- Clevenger, J., Chu, Y., Scheffler, B., and Ozias-Akins, P. (2016). A developmental transcriptome map for allotetraploid arachis hypogaea. *Front. Plant Sci.* 7:1446. doi: 10.3389/fpls.2016.01446
- Cui, G., Duan, L., Jin, B., Qian, J., Xue, Z., Shen, G., et al. (2015). Functional divergence of diterpene syntheses in the medicinal plant *Salvia miltiorrhiza*. *Plant Physiol.* 169, 1607–1618. doi: 10.1104/pp.15.00695
- Dar, A. A., Choudhury, A. R., Kancharla, P. K., and Arumugam, N. (2017). The FAD2 Gene in plants: occurrence, regulation, and role. *Front. Plant Sci.* 8:1789. doi: 10.3389/fpls.2017.01789
- do Vale, T. G., Furtado, E. C., Santos, J. G., and Viana, G. S. B. (2002). Central effects of citral, myrcene and limonene, constituents of essential oil chemotypes

- from *Lippia alba* (Mill.) n.e. Brown. *Phytomed. Int. J. Phytother. Phytopharm.* 9, 709–714. doi: 10.1078/094471102321621304
- Druka, A., Muehlbauer, G., Druka, I., Caldo, R., Baumann, U., Rostoks, N., et al. (2006). An atlas of gene expression from seed to seed through barley development. *Funct. Integr. Genomics* 6, 202–211. doi: 10.1007/s10142-006-0025-4
- Elshafie, H. S., Alibert, L., Amato, M., De Feo, V., and Camele, I. (2018). Chemical composition and antimicrobial activity of Chia (*Salvia hispanica* L.) essential oil. *Eur. Food Res. Technol.* 244, 1675–1682. doi: 10.1007/s00217-018-3080-x
- Ge, X., Chen, H., Wang, H., Shi, A., and Liu, K. (2014). *De novo* assembly and annotation of salvia splendens transcriptome using the illumina platform. *PLoS ONE* 9:e87693. doi: 10.1371/journal.pone.0087693
- Gomez, M. D., Urbez, C., Perez-Amador, M. A., and Carbonell, J. (2011). Characterization of constricted fruit (ctf) mutant uncovers a role for AtMYB117/LOF1 in ovule and fruit development in *Arabidopsis thaliana*. *PLoS ONE* 6:e18760. doi: 10.1371/journal.pone.0018760
- Gómez-Ariza, J., Brambilla, V., Vicentini, G., Landini, M., Cerise, M., Carrera, E., et al. (2019). A transcription factor coordinating internode elongation and photoperiodic signals in rice. *Nat. Plants* 5, 358–362. doi: 10.1038/s41477-019-0401-4
- Grant, E. H., Fujino, T., Beers, E. P., and Brunner, A. M. (2010). Characterization of NAC domain transcription factors implicated in control of vascular cell differentiation in *Arabidopsis* and *Populus*. *Planta* 232, 337–352. doi: 10.1007/s00425-010-1181-2
- Guo, H., Li, L., Ye, H., Yu, X., Algreen, A., and Yin, Y. (2009). Three related receptor-like kinases are required for optimal cell elongation in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7648–7653. doi: 10.1073/pnas.0812346106
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Harkenrider, M., Sharma, R., Vleeschouwer, D. D., Tsao, L., Zhang, X., Chern, M., et al. (2016). Overexpression of rice wall-associated kinase 25 (OsWAK25) alters resistance to bacterial and fungal pathogens. *PLoS ONE* 11:e0147310. doi: 10.1371/journal.pone.0147310
- Haruta, M., Sabat, G., Stecker, K., Minkoff, B. B., and Sussman, M. R. (2014). A peptide hormone and its receptor protein kinase regulate plant cell expansion. *Science* 343, 408–411. doi: 10.1126/science.1244454
- Horiguchi, G., Kim, G.-T., and Tsukaya, H. (2005). The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. *Plant J. Cell Mol. Biol.* 43, 68–78. doi: 10.1111/j.1365-313X.2005.02429.x
- Howe, K. L., Contreras-Moreira, B., De Silva, N., Maslen, G., Akanni, W., Allen, J., et al. (2020). Ensembl genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Res.* 48, D689–D695. doi: 10.1093/nar/gkz890
- Ixtaina, V. Y., Nolasco, S. M., and Tomás, M. C. (2008). Physical properties of Chia (*Salvia hispanica* L.) seeds. *Ind. Crops Prod.* 28, 286–293. doi: 10.1016/j.indcrop.2008.03.009
- Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinform. Oxf. Engl.* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Joyce, B. L., Haug-Baltzell, A. K., Hulvey, J. P., McCarthy, F., Devisetty, U. K., and Lyons, E. (2017). Leveraging CyVerse resources for *de novo* comparative transcriptomics of underserved (non-model) organisms. *J. Vis. Exp.* 123:e55009. doi: 10.3791/55009
- Juergens, U. R. (2014). Anti-inflammatory properties of the monoterpene 1.8-cineole: current evidence for co-medication in inflammatory airway diseases. *Drug Res.* 64, 638–646. doi: 10.1055/s-0034-1372609
- Kagale, S., Nixon, J., Khedikar, Y., Pasha, A., Provart, N. J., Clarke, W. E., et al. (2016). The developmental transcriptome atlas of the biofuel crop *Camelina sativa*. *Plant J. Cell Mol. Biol.* 88, 879–894. doi: 10.1111/tbj.13302
- Kanaoka, M. M., Pillitteri, L. J., Fujii, H., Yoshida, Y., Bogenschutz, N. L., Takabayashi, J., et al. (2008). SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to *Arabidopsis* stomatal differentiation. *Plant Cell* 20, 1775–1785. doi: 10.1105/tpc.108.060848
- Kebrom, T. H., McKinley, B., and Mullet, J. E. (2017). Dynamics of gene expression during development and expansion of vegetative stem internodes of bioenergy sorghum. *Biotechnol. Biofuels* 10:159. doi: 10.1186/s13068-017-0848-3
- Khedher, M. R. B., Khedher, S. B., Chaieb, I., Tounsi, S., and Hammami, M. (2017). Chemical composition and biological activities of *Salvia officinalis* essential oil from Tunisia. *EXCLI J.* 16, 160–173. doi: 10.17179/excli2016-832
- Kim, J. H., Choi, D., and Kende, H. (2003). The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in *Arabidopsis*. *Plant J. Cell Mol. Biol.* 36, 94–104. doi: 10.1046/j.1365-313X.2003.01862.x
- Kobayashi, K., Suzuki, T., Iwata, E., Nakamichi, N., Suzuki, T., Chen, P., et al. (2015). Transcriptional repression by MYB3R proteins regulates plant organ growth. *EMBO J.* 34, 1992–2007. doi: 10.15252/embj.201490899
- Kudapa, H., Garg, V., Chitkineni, A., and Varshney, R. K. (2018). The RNA-Seq-based high resolution gene expression atlas of chickpea (*Cicer arietinum* L.) reveals dynamic spatio-temporal changes associated with growth and development. *Plant Cell Environ.* 41, 2209–2225. doi: 10.1111/pce.13210
- Lampard, G. R., Macalister, C. A., and Bergmann, D. C. (2008). *Arabidopsis* stomatal initiation is controlled by MAPK-mediated regulation of the bHLH SPEECHLESS. *Science* 322, 1113–1116. doi: 10.1126/science.1162263
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lee, B. H., Ko, J.-H., Lee, S., Lee, Y., Pak, J.-H., and Kim, J. H. (2009). The *Arabidopsis* GRF-INTERACTING FACTOR gene family performs an overlapping function in determining organ size as well as multiple developmental properties. *Plant Physiol.* 151, 655–668. doi: 10.1104/pp.109.141838
- Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M. G., et al. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* 29, 1035–1043. doi: 10.1093/bioinformatics/btt337
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R. J., Franklin, L. D., et al. (2010). An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J. Cell Mol. Biol.* 63, 86–99. doi: 10.1111/j.1365-313X.2010.04222.x
- Ma, Y., Yuan, L., Wu, B., Li, X., Chen, S., and Lu, S. (2012). Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *J. Exp. Bot.* 63, 2809–2823. doi: 10.1093/jxb/err466
- MacAlister, C. A., Ohashi-Ito, K., and Bergmann, D. C. (2007). Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. *Nature* 445, 537–540. doi: 10.1038/nature05491
- Manosalva, P. M., Davidson, R. M., Liu, B., Zhu, X., Hulbert, S. H., Leung, H., et al. (2009). A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. *Plant Physiol.* 149, 286–296. doi: 10.1104/pp.108.128348
- Marcinek, K., and Krejpcio, Z. (2017). Chia seeds (*Salvia hispanica*): health promoting properties and therapeutic applications—a review. *Rocz. Panstw. Zakl. Hig.* 68, 123–129.
- Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., et al. (2020). Mass-spectrometry-based draft of the *Arabidopsis* proteome. *Nature* 579, 409–414. doi: 10.1038/s41586-020-2094-2
- MEV: MultiExperiment Viewer (2017). CCCB at Dana-Farber Cancer Institute. Available online at: <https://github.com/dfci-cccb/mev> (accessed December 29, 2017).
- Mikheenko, A., Valin, G., Prjibelski, A., Saveliev, V., and Gurevich, A. (2016). Icarus: visualizer for *de novo* assembly evaluation. *Bioinformatics* 32, 3321–3323. doi: 10.1093/bioinformatics/btw379
- Millar, A. A., and Gubler, F. (2005). The *Arabidopsis* GAMBYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* 17, 705–721. doi: 10.1105/tpc.104.027920

- Mohd Ali, N., Yeap, S. K., Ho, W. Y., Beh, B. K., Tan, S. W., and Tan, S. G. (2012). The Promising Future of Chia, *Salvia hispanica* L. *Biomed. Res. Int.* 2012:171956. doi: 10.1155/2012/171956
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321
- Moss, M., and Oliver, L. (2012). Plasma 1,8-cineole correlates with cognitive performance following exposure to rosemary essential oil aroma. *Ther. Adv. Psychopharmacol.* 2, 103–113. doi: 10.1177/2045125312436573
- Mount, D. W. (2007). Using the Basic Local Alignment Search Tool (BLAST). *Cold Spring Harb. Protoc.* 2007:pdb.top17. doi: 10.1101/pdb.top17
- Muñoz, L. A., Cobos, A., Diaz, O., and Aguilera, J. M. (2013). Chia seed (*Salvia hispanica*): an ancient grain and a new functional food. *Food Rev. Int.* 29, 394–408. doi: 10.1080/87559129.2013.818014
- Naithani, S., Nonogaki, H., and Jaiswal, P. (2017). “Exploring crossroads between seed development and stress response,” in *Mechanism of Plant Hormone Signaling Under Stress*, ed G. K. Pandey. doi: 10.1002/9781118889022.ch32
- najoshi/sickle. *GitHub*. Available online at: <https://github.com/najoshi/sickle> (accessed June 6, 2014).
- Nakashima, K., Shinwari, Z. K., Sakuma, Y., Seki, M., Miura, S., Shinozaki, K., et al. (2000). Organization and expression of two Arabidopsis DREB2 genes encoding DRE-binding proteins involved in dehydration- and high-salinity-responsive gene expression. *Plant Mol. Biol.* 42, 657–665. doi: 10.1023/A:1006321900483
- Oliva, M. E., Ferreira, M. R., Chicco, A., and Lombardo, Y. B. (2013). Dietary Salba (*Salvia hispanica* L) seed rich in α -linolenic acid improves adipose tissue dysfunction and the altered skeletal muscle glucose and lipid metabolism in dyslipidemic insulin-resistant rats. *Prostaglandins Leukot. Essent. Fat. Acids* 89, 279–289. doi: 10.1016/j.plefa.2013.09.010
- Papatheodorou, I., Fonseca, N. A., Keays, M., Tang, Y. A., Barrera, E., Bazant, W., et al. (2018). Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.* 46, D246–D251. doi: 10.1093/nar/gkx1158
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197
- Peiretti, P. G., and Gai, F. (2009). Fatty acid and nutritive quality of Chia (*Salvia hispanica* L.) seeds and plant during growth. *Anim. Feed Sci. Technol.* 148, 267–275. doi: 10.1016/j.anifeeds.2008.04.006
- Peláez, P., Orona-Tamayo, D., Montes-Hernández, S., Valverde, M. E., Paredes-López, O., and Cibrián-Jaramillo, A. (2019). Comparative transcriptome analysis of cultivated and wild seeds of *Salvia hispanica* (Chia). *Sci. Rep.* 9:9761. doi: 10.1038/s41598-019-45895-5
- Reeves, P. H., Ellis, C. M., Ploense, S. E., Wu, M.-F., Yadav, V., Tholl, D., et al. (2012). A regulatory network for coordinated flower maturation. *PLoS Genet.* 8:e1002506. doi: 10.1371/journal.pgen.1002506
- Reyes-Caudillo, E., Tecante, A., and Valdivia-López, M. A. (2008). Dietary fibre content and antioxidant activity of phenolic compounds present in Mexican Chia (*Salvia hispanica* L.) seeds. *Food Chem.* 107, 656–663. doi: 10.1016/j.foodchem.2007.08.062
- Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2002). DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.* 290, 998–1009. doi: 10.1006/bbrc.2001.6299
- Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. doi: 10.1093/bioinformatics/bts094
- Sekhon, R. S., Lin, H., Childs, K. L., Hansey, C. N., Buell, C. R., de Leon, N., et al. (2011). Genome-wide atlas of transcription during maize development. *Plant J. Cell Mol. Biol.* 66, 553–563. doi: 10.1111/j.1365-313X.2011.04527.x
- Severin, A. J., Woody, J. L., Bolon, Y.-T., Joseph, B., Diers, B. W., Farmer, A. D., et al. (2010). RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol.* 10:160. doi: 10.1186/1471-2229-10-160
- Shpak, E. D., McAbee, J. M., Pillitteri, L. J., and Torii, K. U. (2005). Stomatal patterning and differentiation by synergistic interactions of receptor kinases. *Science* 309, 290–293. doi: 10.1126/science.1109710
- Sibout, R., Eudes, A., Mouille, G., Pollet, B., Lapierre, C., Jouanin, L., et al. (2005). CINNAMYL ALCOHOL DEHYDROGENASE-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of Arabidopsis. *Plant Cell* 17, 2059–2076. doi: 10.1105/tpc.105.030767
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Smith-Unna, R., Bournsnel, C., Patro, R., Hibberd, J., and Kelly, S. (2016). TransRate: reference free quality assessment of *de novo* transcriptome assemblies. *Genome Res.* 26, 1134–1144. doi: 10.1101/gr.196469.115
- Sreedhar, R. V., Kumari, P., Rupwate, S. D., Rajasekharan, R., and Srinivasan, M. (2015). Exploring triacylglycerol biosynthetic pathway in developing seeds of chia (*Salvia hispanica* L.): a transcriptomic approach. *PLoS ONE* 10:e0123580. doi: 10.1371/journal.pone.0123580
- Stelpflug, S. C., Sekhon, R. S., Vaillancourt, B., Hirsch, C. N., Buell, C. R., de Leon, N., et al. (2016). An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome* 9, 1–16. doi: 10.3835/plantgenome2015.04.0025
- Tello-Ruiz, M. K., Naithani, S., Gupta, P., Olson, A., Wei, S., Preece, J., et al. (2021). Gramene 2021: harnessing the power of comparative genomics and pathways for plant research. *Nucleic Acids Res.* 49, D1452–D1463. doi: 10.1093/nar/gkaa979
- Temnykh, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 11, 1441–1452. doi: 10.1101/gr.184001
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382
- TransDecoder. *Find Coding Regions Within Transcripts*. Available online at: https://transdecoder.github.io/#incl_homology (accessed January 23, 2017).
- Trikka, F. A., Nikolaidis, A., Ignea, C., Tsaballa, A., Tziveleka, L.-A., Ioannou, E., et al. (2015). Combined metabolome and transcriptome profiling provides new insights into diterpene biosynthesis in *S. pomifera* glandular trichomes. *BMC Genomics* 16:935. doi: 10.1186/s12864-015-2147-3
- Tzfadia, O., Diels, T., De Meyer, S., Vandepoele, K., Aharoni, A., and Van de Peer, Y. (2015). CoExpNetViz: comparative Co-expression networks construction and visualization tool. *Front. Plant Sci.* 6:1194. doi: 10.3389/fpls.2015.01194
- Ullah, R., Nadeem, M., Khalique, A., Imran, M., Mehmood, S., Javid, A., et al. (2016). Nutritional and therapeutic perspectives of Chia (*Salvia hispanica* L.): a review. *J. Food Sci. Technol.* 53, 1750–1758. doi: 10.1007/s13197-015-1967-0
- Valdivia-López, M. Á., and Tecante, A. (2015). Chia (*Salvia hispanica*): a review of native mexican seed and its nutritional and functional properties. *Adv. Food Nutr. Res.* 75, 53–75. doi: 10.1016/bs.afnr.2015.06.002
- Vining, K. J., Johnson, S. R., Ahkami, A., Lange, I., Parrish, A. N., Trapp, S. C., et al. (2017). Draft genome sequence of mentha longifolia and development of resources for mint cultivar improvement. *Mol. Plant* 10, 323–339. doi: 10.1016/j.molp.2016.10.018
- Vuksan, V., Choleva, L., Jovanovski, E., Jenkins, A. L., Au-Yeung, F., Dias, A. G., et al. (2017a). Comparison of flax (*Linum usitatissimum*) and Salba-Chia (*Salvia hispanica* L.) seeds on postprandial glycemia and satiety in healthy individuals: a randomized, controlled, crossover study. *Eur. J. Clin. Nutr.* 71, 234–238. doi: 10.1038/ejcn.2016.148
- Vuksan, V., Jenkins, A. L., Brissette, C., Choleva, L., Jovanovski, E., Gibbs, A. L., et al. (2017b). Salba-Chia (*Salvia hispanica* L.) in the treatment of overweight and obese patients with type 2 diabetes: a double-blind randomized controlled trial. *Nutr. Metab. Cardiovasc. Dis.* 27, 138–146. doi: 10.1016/j.numecd.2016.11.124
- Vuksan, V., Jenkins, A. L., Dias, A. G., Lee, A. S., Jovanovski, E., Rogovik, A. L., et al. (2010). Reduction in postprandial glucose excursion and prolongation of satiety: possible explanation of the long-term effects of whole grain Salba (*Salvia hispanica* L.). *Eur. J. Clin. Nutr.* 64, 436–438. doi: 10.1038/ejcn.2009.159
- Vuksan, V., Whitham, D., Sievenpiper, J. L., Jenkins, A. L., Rogovik, A. L., Bazinet, R. P., et al. (2007). Supplementation of conventional therapy with the novel grain Salba (*Salvia hispanica* L.) improves major and emerging cardiovascular risk factors in type 2 diabetes: results of a randomized controlled trial. *Diabetes Care* 30, 2804–2810. doi: 10.2337/dc07-1144

- Wang, H., Niu, H., Liang, M., Zhai, Y., Huang, W., Ding, Q., et al. (2019). A wall-associated kinase gene CaWAKL20 from pepper negatively modulates plant thermotolerance by reducing the expression of ABA-responsive genes. *Front. Plant Sci.* 10:591. doi: 10.3389/fpls.2019.00591
- Wang, T., Chen, X., Zhu, F., Li, H., Li, L., Yang, Q., et al. (2013). Characterization of peanut germin-like proteins, AhGLPs in plant development and defense. *PLoS ONE* 8:e61722. doi: 10.1371/journal.pone.0061722
- Wenping, H., Yuan, Z., Jie, S., Lijun, Z., and Zhezhi, W. (2011). *De novo* transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. *Genomics* 98, 272–279. doi: 10.1016/j.ygeno.2011.03.012
- Wimberley, J., Cahill, J., and Atamian, H. S. (2020). *De novo* sequencing and analysis of *salvia hispanica* tissue-specific transcriptome and identification of genes involved in terpenoid biosynthesis. *Plants* 9:405. doi: 10.3390/plants9030405
- Xue, Y., Chen, B., Win, A. N., Fu, C., Lian, J., Liu, X., et al. (2018). Omega-3 fatty acid desaturase gene family from two ω -3 sources, *Salvia hispanica* and *Perilla frutescens*: cloning, characterization, and expression. *PLoS ONE* 13:e0191432. doi: 10.1371/journal.pone.0191432
- Yang, C., Xu, Z., Song, J., Conner, K., Vizcay Barrena, G., and Wilson, Z. A. (2007). *Arabidopsis* MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. *Plant Cell* 19, 534–548. doi: 10.1105/tpc.106.046391
- Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107
- Zhang, B., Li, P., Su, T., Li, P., Xin, X., Wang, W., et al. (2020). Comprehensive analysis of wall-associated kinase genes and their expression under abiotic and biotic stress in chinese cabbage (*brassica rapa* ssp. *pekinensis*). *J. Plant Growth Regul.* 39, 72–86. doi: 10.1007/s00344-019-09964-3
- Zhang, H., Miao, H., Wang, L., Qu, L., Liu, H., Wang, Q., et al. (2013). Genome sequencing of the important oilseed crop *Sesamum indicum*L. *Genome Biol.* 14:401. doi: 10.1186/gb-2013-14-1-401
- Zhang, J., Liu, H., Sun, J., Li, B., Zhu, Q., Chen, S., et al. (2012). *Arabidopsis* fatty acid desaturase FAD2 is required for salt tolerance during seed germination and early seedling growth. *PLoS ONE* 7:e30355. doi: 10.1371/journal.pone.0030355
- Zhong, R., Lee, C., Zhou, J., McCarthy, R. L., and Ye, Z.-H. (2008). A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *Plant Cell* 20, 2763–2782. doi: 10.1105/tpc.108.061325
- Zimmermann, G., Bäumlein, H., Mock, H.-P., Himmelbach, A., and Schweizer, P. (2006). The multigene family encoding germin-like proteins of barley. Regulation and function in Basal host resistance. *Plant Physiol.* 142, 181–192. doi: 10.1104/pp.106.083824
- Zimmermann, I. M., Heim, M. A., Weisshaar, B., and Uhrig, J. F. (2004). Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J. Cell Mol. Biol.* 40, 22–34. doi: 10.1111/j.1365-313X.2004.02183.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Gupta, Geniza, Naithani, Phillips, Haq and Jaiswal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.