G3 Genes | Genomes | Genetics

INVESTIGATION

Tissue-Specific Transcriptome Analysis Reveals Candidate Genes for Terpenoid and Phenylpropanoid Metabolism in the Medicinal Plant *Ferula assafoetida*

Hajar Amini,*,† Mohammad Reza Naghavi,†,1 Tong Shen,‡ Yanhong Wang,§ Jaber Nasiri,† Ikhlas A. Khan,§ Oliver Fiehn,‡, Philipp Zerbe,* and Julin N. Maloof*,1**

*Department of Plant Biology and ‡West Coast Metabolomics Center, University of California, Davis, CA 95616, †Department of Agronomy and Plant Breeding, Agricultural and Natural Resources College, University of Tehran, Iran, 77871-31587, §National Center for Natural Products Research, Research Institute of Pharmaceutical Sciences, University of Mississippi, Oxford, MS 38677, and **Department of Biochemistry, Faculty of Sciences, King Abdulaziz University, Jeddah, Saudi-Arabia

ORCID IDs: 0000-0002-6261-8928 (O.F.); 0000-0001-5163-9523 (P.Z.); 0000-0002-9623-2599 (J.N.M.)

ABSTRACT *Ferula assafoetida* is a medicinal plant of the Apiaceae family that has traditionally been used for its therapeutic value. Particularly, terpenoid and phenylpropanoid metabolites, major components of the root-derived oleo-gum-resin, exhibit anti-inflammatory and cytotoxic activities, thus offering a resource for potential therapeutic lead compounds. However, genes and enzymes for terpenoid and coumarin-type phenylpropanoid metabolism have thus far remained uncharacterized in *F. assafoetida*. Comparative *de novo* transcriptome analysis of roots, leaves, stems, and flowers was combined with computational annotation to identify candidate genes with probable roles in terpenoid and coumarin biosynthesis. Gene network analysis showed a high abundance of predicted terpenoid- and phenylpropanoid-metabolic pathway genes in flowers. These findings offer a deeper insight into natural product biosynthesis in *F. assafoetida* and provide genomic resources for exploiting the medicinal potential of this rare plant.

KEYWORDS
RNAseq
coumarin-type phenylpropanoids
terpenoid
medicinal plant

Ferula assafoetida L. (Apiaceae) is a herbaceous, monoecious and perennial plant indigenous to Kashmir, Afghanistan, and Iran (Iranshahy and Iranshahi 2011). *F. assafoetida* is predominantly valued for its medicinal uses as an important source of oleo-gum-resin, called asafoetida, that is obtained from the exudates of the tap roots (Kavoosi and Rowshan 2013). Asafoetida has broad therapeutic properties, for example for the treatment of inflammations, neurological and digestive disorders, rheumatism, headache, arthritis, and dizziness (Iranshahy and Iranshahi 2011). The oleo-gum-resin consists of three main fractions, including resin (40–64%), gum (25%) and essential oil (10–17%) (Amalraj and Gopi 2017). Phenylpropanoids (especially coumarins-related compounds) and terpenoids are the major constituents of the resin (Amalraj and Gopi 2017), whereas the essential oil is mostly comprised of sulfur-containing compounds, as well as volatile mono- and sesqui-terpenoids (Divya *et al.* 2014). Among these metabolites, sesquiterpene coumarins are of particular importance, due to their extensive and promising biological properties (Iranshahi *et al.* 2008). Sesquiterpene coumarins, which contain a coumarin or 1-benzopyran-2-one group joint with a sesquiterpene scaffold, are almost exclusively found in the genus *Ferula* and accumulate mainly in the roots of the plant (Curini *et al.* 2006).

In plants, terpenoid metabolites are derived from two isomeric 5-carbon precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are formed via the plastidial methyl-erythritol-5-phosphate (MEP) and the cytosolic mevalonate (MEV)

doi: https://doi.org/10.1534/g3.118.200852
Manuscript received October 31, 2018; accepted for publication January 11, 2019; published Early Online January 24, 2019.

Supplemental material available at Figshare: https://doi.org/10.25387/g3.7609223.
[1]Corresponding authors: Department of Plant Biology, 1 Shields Ave, University of California, Davis, CA 95616, E-mail: jnmaloof@ucdavis.edu, Department of Agronomy and Plant Breeding, Daneshkadeh Ave, Agricultural and Natural Resources College, University of Tehran, Tehran, Iran, 77871-31587, E-mail: mnaghavi@ut.ac.ir

pathway (Chen *et al.* 2011; Falara and Pichersky 2012). Prenyltransferase-catalyzed condensation of IPP and DMAPP units into prenyl diphosphate intermediates of different chain length provides central precursors that are further converted by species-specific enzymes families of terpene synthases (TPSs) and cytochrome P450 monooxygenases (P450s) to give rise to the chemical diversity of plant terpenoids (Chen *et al.* 2011; Pateraki *et al.* 2015) (Supplementary Figure S1A). In addition to terpenoids, flavonoids (especially luteolin) and coumarin-related compounds (especially umbelliferone) are enriched in *F. assafoetida* (Pangarova and Zapesochnaya 1973; Amalraj and Gopi 2017). These metabolites are derived from p-coumaroyl-CoA formed via the plastidial shikimate pathway (Naoumkina *et al.* 2010; Vogt 2010). Branching from this core intermediate, distinct 2-oxoglutarate-dependent dioxygenases, feruloyl-CoA-6'-hydroxylases (F6'H) and coumaroyl-CoA-2'-hydroxylases (C2'H), facilitate the key steps in the biosynthesis of umbelliferone and other coumarins (Matsumoto *et al.* 2012; Vialart *et al.* 2012; Tohge *et al.* 2013). By contrast, flavonoid biosynthesis requires the conversion of p-coumaroyl-CoA by a chalcone synthase (CHS) and a chalcone isomerase (CHI), followed by various possible functional modifications of the resulting naringenin intermediate (Naoumkina *et al.* 2010; Vogt 2010) (Supplementary Figure S1B).

Rapid advances in genomics and biochemical technologies have enabled a deeper investigation of metabolic pathways in range of medicinal and other non-model plants (Xiao *et al.* 2013; Zerbe *et al.* 2013; Kitaoka *et al.* 2015; Wurtzel and Kutchan 2016). Among members of the Apiaceae, genomics-enabled gene discovery in carrot (*Daucus carota*) was utilized to identify flavonoid and isoprenoid pathway genes (Iorizzo *et al.* 2016) and enabled the functional characterization of two carrot terpene synthase genes, the sesquiterpene synthase *DcTPS1* and the monoterpene synthase *DcTPS2* (Yahyaa *et al.* 2015).

In this study, we employed comparative *de novo* transcriptome analyses and computational gene annotation to identify candidate genes with probable roles in the biosynthesis of bioactive terpenoid and coumarin-type phenylpropanoid metabolites in *F. assafoetida*.

MATERIAL AND METHODS

Plant material

Roots, stems, flowers, and leaves of *F. assafoetida* were collected from the Molla Ahmad Mountains, Isfahan province, Iran, at an altitude of 2250 meters (53°35′E and 32°15′N). Samples were collected from three separate plants, which were used as biological replicates. The identity of the harvested plants was verified by the Iranian Biological Resource Center (IBRC).

Metabolite analysis

Roots, stems, flowers, and leaves were air-dried in the shade at room temperature. Terpenoid and sesquiterpene-coumarin compounds present in the oleo-gum-resin were extracted from the essential oil. In the context of this study, essential oils were prepared by grinding 20 grams of plant organs (roots, stems, flowers, and leaves) to a fine powder. Essential oils were then isolated through hydro-distillation for 5 hr, using a Clevenger type apparatus. The distilled oils were dried over anhydrous sodium sulfate and after filtration stored at 4°. Terpenoid analysis was then performed via GC-MS analysis as described in the Supplementary Methods. Product identification was achieved by comparison to reference mass spectra and retention indices (RI) available through the US National Institute of Standards and Technology (NIST, USA), WILEY 1996 Ed. mass spectral library, as well as an in-house library.

Umbelliprenin, umbelliferone, and luteolin were quantified in different organs and oleo-gum-resins of *F. assafoetida* by ultra-high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UHPLC-QToF-MS) at the National Center for Natural Products Research at the University of Mississippi as outlined in the Supplementary Methods.

RNAseq library preparation and pre-processing of short reads

RNA was extracted using BioZOL total RNA extraction kit (BioFlux, Japan) as detailed in the manufacturer's instructions. For removing polyphenol and polysaccharide content in different organs, especially roots, an additional purification step was applied as follows: The resulting pellets were dissolved in Diethyl Pyrocarbonate (DEPC)-treated water and extracted once with phenol-chloroform (1:1) and then with chloroform. The aqueous solution was transferred into 2 new tubes, and 3M sodium acetate (pH 5.2) with 5M NaCl and 0.6 volume of cold isopropanol was added. This solution was mixed and then stored at -20° for one hour. Next, the solution was centrifuged at 13,000 rpm for 10 min at 4°. The pellet was collected after centrifugation by discarding the upper aqueous phase. The pellet was washed twice with 75% ethanol by re-suspending the pellet and centrifuging at 10,000 rpm for 10 min at 4° at each of the wash steps. The ethanol was allowed to evaporate, and the pellet was resuspended in 40 μl of DEPC-treated water. Furthermore, we removed any contaminating DNA using RNase-free DNase I (Thermo Fisher Scientific Inc). Quality and quantity of RNA were assessed by separation of RNA by gel electrophoresis on a 1% agarose gel, NanoDrop (ND-1000) and Agilent 2100 Bioanalyzer. RNA samples with RNA integrity number (RIN) values >8.0 were selected for constructing libraries. RNA libraries were obtained according to Breath Adapter Directional sequencing (BrAD-seq) protocol (Townsley *et al.* 2015) with shotgun (SHO) type strand-specific libraries for sending to the DNA Technologies Core at UC Davis. RNA sequencing was performed taking steps for mRNA Fragmentation, 3-prime Adapter Priming and cDNA Synthesis, 5-prime Duplex Breath Capture Adapter Addition (Strand Specific) and enrichment and adapter extension. The paired-end sequencing with read length of 150 bp was performed on an Illumina HiSeq 4000 platform by the DNA Technologies Core at the UC Davis Genome Center. Another set of RNA samples were sent to the Beijing Genomics Institute (BGI) and Novogene for library preparation and sequencing. After sequencing, raw reads were separated by barcode and filtered by quality using the HiSeq 4000 software CASAVA V1.8. The read quality before and after quality control with Trimmomatic was tested with FastQC quality assessment (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and results were collected from all samples into a single report for easy comparison with MultiQC (Ewels *et al.* 2016). The parameters used in Trimmomatic V0.33 (Bolger *et al.* 2014) for trimming and cropping the FASTQ data as well as removing adapters was set as follows: ILLUMINACLIP:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, MINLEN:120. The summary of the RNAseq reads after trimming, cropping, and adapter removal is shown in Supplementary Table S2.

De novo transcriptome assembly and evaluation

We utilized four different *de novo* transcriptome assembly pipelines to ensure that a high quality reference transcriptome was assembled (Supplemental Table S3). These four different *de novo* transcriptome assembly pipelines are described in more detail in the Supplementary Methods.

Reads from all organs were combined and assembled with Trinity v2.4.0 with kmer 25 (Grabherr *et al.* 2011) and Oases v0.2.06 with kmer 25, 31,37, 43, 49 (Zerbino and Birney 2008; Schulz *et al.* 2012). Drap v1.91 was applied as a post processing step to compact and correct each assembled transcriptome (Cabau *et al.* 2017). In addition to these methods, we used Khmer v2.0 tools (Crusoe *et al.* 2015) to apply variable kmer coverage abundance trimming to the reads prior to Trinity assembly. This reduces the computational cost of assembly without negatively affecting the quality of the assembly. Several different approaches were utilized to assess the quality of each assembled transcriptome. First, we investigated the length distribution of the transcripts produced by the different pipelines (Supplementary Figure S2A). Second, we mapped the reads from each sample to each assembled transcriptome with STAR 2.5.2b to determine the percentage of reads that mapped to each assembly (Supplementary Figure S2B) (Dobin *et al.* 2013). Finally, we assessed the completeness of each assembled transcriptome in terms of expected genes with BUSCO v3 (Simão *et al.* 2015; Waterhouse *et al.* 2018) and using "Plant set (Embryophyta *odb9*)" as a database of BUSCO group with 1440 genes (Supplementary Table S4).

Transcriptome annotation

We first annotated the assembled transcriptome using the dammit tool (Scott 2016). In this analysis, dammit searched the well-known annotated protein databases, for example, Pfam (Finn *et al.* 2015), Rfam (Kalvari *et al.* 2017), OrthoDB (Zdobnov *et al.* 2016), BUSCO, and UniRef (Suzek *et al.* 2014) for significant matches against the *F. assafoetida* assembled transcriptomes using blast (E-value cutoff $<1e^{-5}$). A large fraction of the assembled transcripts was assigned successfully to known annotated protein from other plants. Out of the 60,134 assembled transcripts, dammit was able to map 54,129 (>90%) to known genes using blast with an E-value <0.00001 (for each hit).

In addition, the most likely protein sequence was identified by using TransDecoder (http://transdecoder.github.io) to find the longest open reading frame. Furthermore, sequences were initially annotated by comparing *F. assafoetida* protein sequences against the Arabidopsis protein sequence database (TAIR10) (n = 35,386). We were able to successfully find a significant hit (E-value $<1e^{-10}$) to 30,344 (>85%) of the Arabidopsis protein sequence, which covered 34,920 (58%) of the *F. assafoetida* assembled transcripts. We also used blastx to compare the assembled transcriptome against all RefSeq plants database (>10,710 plants) (Pruitt *et al.* 2006). Our observations indicated that a significant majority of our assembled contigs (>77%) had the best hit with *D. carota* (Supplementary Figure S3). This is expected given the close relationship of *F. assafoetida* and *D. carota* (Ahmed *et al.* 2005).

Gene ontology (GO) analysis was performed on the whole assembled transcriptome using Blast2GO v1.3.3 (Conesa *et al.* 2005; Götz *et al.* 2008). Blast2GO allowed us to identify similarity of the *F. assafoetida* sequences to GenBank non-redundant proteins (Nr) and Swiss-Prot databases using blastx (E-value $<1e^{-5}$ with the number of hits limited to a maximum of twenty). This was followed by InterProScan search, mapping, and annotation using standard parameters from Blast2GO. After obtaining GO annotation for every transcript, WEGO software (Ye *et al.* 2006) was then used to simplify the output for producing combined graphs for molecular function, cellular process, and biological process. (Supplementary Figure S4). Based on Gene Ontology analyses, a total of 31,524 (52.42%) transcripts had one or more terms assigned.

Phylogenetic analysis

Protein sequence alignments were generated using CLC bio software (www.Qiagen.com) followed by manual curation. Maximum-likelihood phylogenetic analyses were performed using PhyML version 3.0.1 beta (Anisimova *et al.* 2011) with four rate substitution categories, LG substitution model, BIONJ starting tree and 500 bootstrap repetitions.

Tissue-specific differential expression analysis

To identify candidate genes for terpenoid and phenylpropanoid metabolism in *F. assafoetida*, we need to compare expression level of genes across different organs. So, the differential expression analysis was utilized to determine how gene expression of target processes differ between different organs.

To determine patterns of gene expression across the different organs, RNA abundance in the assembled transcriptome was quantified using Kallisto (Bray *et al.* 2016), then the differential expression analysis was calculated from 12 samples from UC Davis facility (3 samples from 4 different organs) using the edgeR package in the R statistical environment (FDR <0.05) (Robinson *et al.* 2010b; R Core Team 2016). Sample libraries were normalized by calculating the effective library size and normalization factor using the TMM method on counts data (Robinson and Oshlack 2010a). We then identified genes differentially expressed among plant organs using a generalized linear model (glm) in edgeR and multiple-testing correction via the Benjamini and Hochberg (BH) procedure (Benjamini and Hochberg 1995). The Venn diagram of differential expressed genes of pairwise comparison of different organs was done using Intervene tool (Khan and Mathelier 2017). We provide additional information regarding the experimental design of differential expression analysis in the Supplementary Materials.

Over-representation analysis of differentially expressed genes

To find significantly enriched GO terms, over-representation analysis (ORA) was done by GOseq package (Young *et al.* 2010), followed by (BH) multiple testing correction to achieve an experiment-wise threshold of $P < 0.05$. Moreover, we investigated the KEGG database by using BlastKOALA (Kanehisa *et al.* 2016) to assign KEGG Orthology (KO) to each transcript. Associated plant-related KEGG pathway ids were obtained from the KO by using KEGGREST Bioconductor package (Tenenbaum 2018). Over-representation analysis of KEGG pathways was also conducted by GOseq package at cut off value p-adjust <0.05 (Young *et al.* 2010) in the R statistical environment (R Core Team 2016). The enriched pathways were visualized with the Pathview Bioconductor package (Luo and Brouwer 2013).

Gene network analysis (WGCNA)

To find genetic modules that were highly co-expressed across different organs, we performed a weighted gene co-expression network analysis using the WGCNA package v1.63 in R/Bioconductor (Langfelder and Horvath 2008). First, pairwise gene co-expression was calculated from the 12 samples from UC Davis facility (3 samples from 4 different organs) to avoid the possibility of a batch effect that could occur if we included samples sequenced at other facilities.

We further investigated the optimal power for constructing the gene co-expression as indicated in WGCNA best practices and picked the value 12 (Supplementary Figure S5) as a soft power. The network was constructed by setting the type to signed hybrid, minModuleSize to 30, dissimilarity threshold to 0.2, and deepslit to 2.

Data availability

All transcriptome data were submitted to the NCBI sequence read archive (SRA) with accession number (PRJNA476150) and Temporary Submission ID (SUB4158602). All R scripts for this paper are available

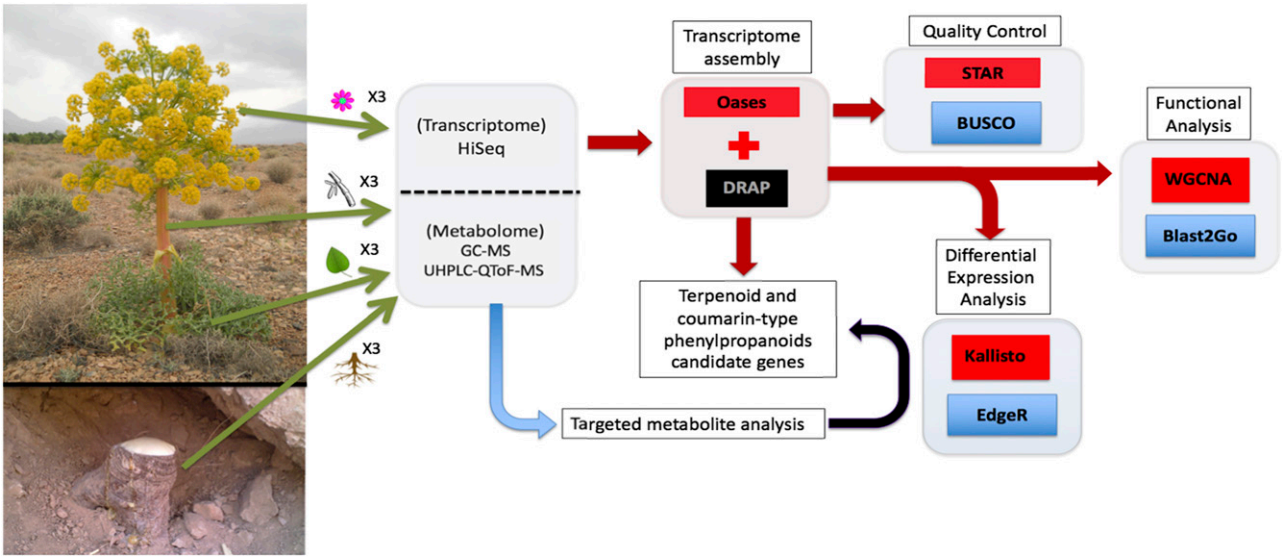

Figure 1 Workflow for transcriptome and metabolite analysis of different organs of *F. assafoetida*. Shown is a schematic overview of the transcriptomics analysis performed on different organs (flowers, leaves, stems, and roots) harvested from three wild *F. assafoetida* plants (green arrows) and used for transcriptome (red) and targeted metabolite analysis (blue). Black arrows highlight the correlation of *de novo* transcriptome and metabolite analysis to identify candidate genes with possible roles in terpenoid and coumarin-type phenylpropanoid metabolism.

at https://github.com/MaloofLab/Amini-G3-2019-Ferula_RNAseq_Analysis. Supplemental material available at Figshare: https://doi.org/10.25387/g3.7609223.

RESULTS AND DISCUSSION

To identify the tissue-specific abundance of terpenoid and coumarin-type phenylpropanoid biosynthetic genes we employed correlation studies of gene expression and metabolite abundance across select plant organs (Figure 1).

Abundance of major oleo-resin metabolites

Coumarin- and flavonoid-type phenylpropanoids, as well as terpenoids are major bioactive constituents of *F. assafoetida*. To investigate the distribution of these metabolites between different organs, we measured the abundance of select terpenoids, flavonoids (luteolin) and coumarins (umbelliferone) using liquid and gas chromatography-mass spectrometry (LC/GC-MS). Among the identified terpenoid metabolites, β-pinene, α-pinene, γ-elemene, β-maaliene, and α/β-eudesmol were the most abundant. Furthermore, terpenoids differed in their abundance in different organs and accumulated at the highest level in roots containing near equal amounts of mono- and sesqui-terpenoids that formed 54% of the resin-derived essential oil (Supplementary Figure S6). This is consistent with previous studies in *D. carota* that suggested that terpenoid metabolism differs substantially between organs (Habegger and Schnitzler 2000).

The coumarin umbelliferone and the flavone luteolin are commonly occurring phenylpropanoid metabolites in members of the Apiaceae family and serve as a precursor for a range of specialized metabolites, including pyrano-coumarins and furano-coumarins (Luo *et al.* 2017; Yao *et al.* 2017). Therefore, we quantified umbelliferone and luteolin across the select plant organs (Supplementary Figure S7A and S7B) and isolated oleo-gum-resin fractions (Supplementary Figure S7C). Umbelliferone and luteolin were most abundant in roots (1.95 ± 0.8 μg/g) and flowers (99.72 ± 40.75 μg/g), respectively (Supplementary Figure S7A and S7B).

De novo transcriptome assembly and evaluation

Since no reference genome is available for the genus *Ferula*, we performed *de novo* transcriptome assembly to obtain a reference transcriptome assembly as described in the Supplementary Materials. The Oases_DRAP *de novo* transcriptome assembly was used for further analysis.

Transcriptome annotation

To gain additional insight into biosynthetic pathways all transcripts were queried against the KEGG database with blastKOALA. We found that 43,453 of the 60,134 assembled transcripts were assigned successfully to the KEGG database. We next queried the generated transcriptome data for key genes in terpenoid and phenylpropanoid metabolism. Using homology searches against manually curated protein databases of key gene families with an E-value cut-off of $1e^{-75}$ (Zerbe *et al.* 2013), we identified 27 candidates for MEV and MEP pathway genes, 32 transcripts with significant matches to terpene synthase (TPS) and triterpene synthase (TTS) genes, and 245 transcripts representing putative P450s. In addition, 142 transcripts with significant matches to phenylpropanoid pathway genes were identified.

Phylogenetic analysis of transcripts with predicted functions in terpenoid and phenylpropanoid metabolism

Of the identified *F. assafoetida* transcripts, 16 TPS- and TTS-like sequences, as well as 23 putative phenylpropanoid-metabolic enzymes, which represented full-length sequences with the highest similarity to known enzymes were selected for further phylogenetic analysis. To infer possible functions, phylogenetic analysis of these enzyme candidates was performed in comparison to previously reported protein sequences of related Apiaceae species (including *Daucus carota* and *Thapsia garganica*), as well as proteins from Asteraceae and other dicot species that represent key pathway reactions. For clarity, all gene candidates further investigated here have been assigned gene designations based on their predicted function, using common abbreviations for terpenoid- and phenylpropanoid-metabolic enzymes. The corresponding transcript identifiers are given in Supplementary Table S5 and S6.

Of the identified TPS-related genes, nine candidates were most closely related to members of the TTS family, including predicted cycloartenol

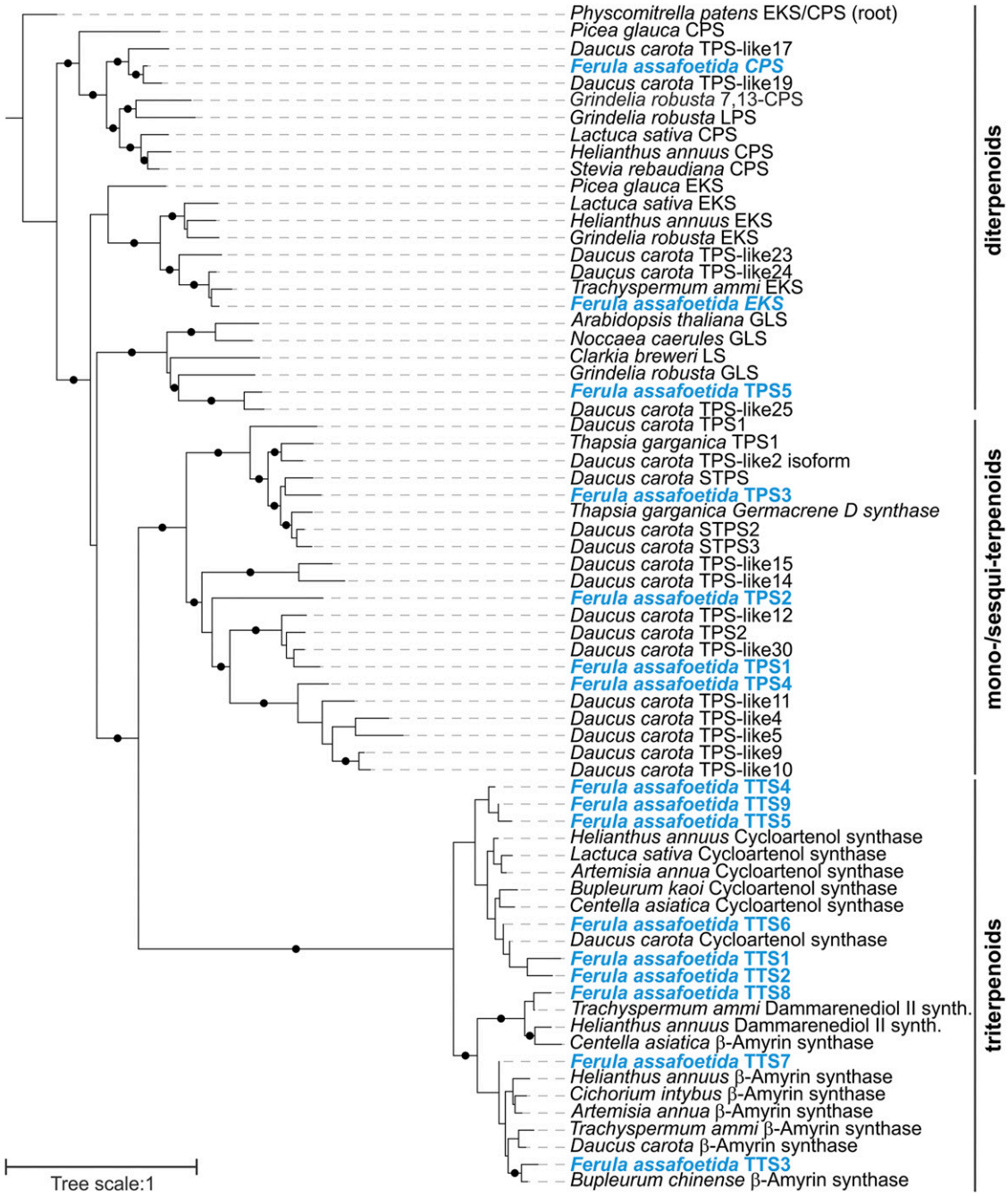

Figure 2 Maximum-likelihood phylogenetic tree of terpene synthase (TPS) and triterpene synthase (TTS) candidates identified in *F. assafoetida* as compared to known enzymes from related plant species. The *Physcomitrella patens ent*-kaurene synthase/copalyl diphosphate synthase (EKS/CPS) was used to root the tree. Branches with bootstrap support of >80% (500 repetitions) are highlighted with black dots.

synthases (TTS1-2, 4-6 and 9) with possible roles in sterol biosynthesis, and β-amyrin synthases (TTS3, 7 and 8) putatively involved in the formation of specialized triterpenoid metabolites (Figure 2). In addition, seven candidates were placed within the TPS family. Among these, *F. assafoetida* CPS and EKS clustered with known *ent*-copalyl diphosphate synthases (CPS) and *ent*-kaurene synthases (EKS) with widely conserved functions in gibberellin phytohormone biosynthesis (Peters 2010; Zerbe and Bohlmann 2015), suggesting a similar function. In addition, *F. assafoetida* TPS7 clustered with known geranyllinalool synthases involved in the biosynthesis of defensive homoterpene metabolites (Herde *et al.* 2008; Falara *et al.* 2014; Richter *et al.* 2016). The remaining TPS candidates were placed within the group of mono- and sesqui-TPSs including characterized and predicted TPSs of the close relative *D. carota* (Yahyaa *et al.* 2015; Yahyaa *et al.* 2018). Combined with their phylogenetic relationships, presence of plastidial transit peptides and a characteristic RRX8W motif (Chen *et al.* 2011) suggested a monoterpene synthase function for TPS1 and TPS4, whereas absence of these features indicated a sesquiterpene synthase activity for *F. assafoetida* TPS2 and TPS3. However, an only distant relationship to currently known Apiaceae TPSs, such as the *D. carota* β-caryophyllene synthase DcTPS1 and the

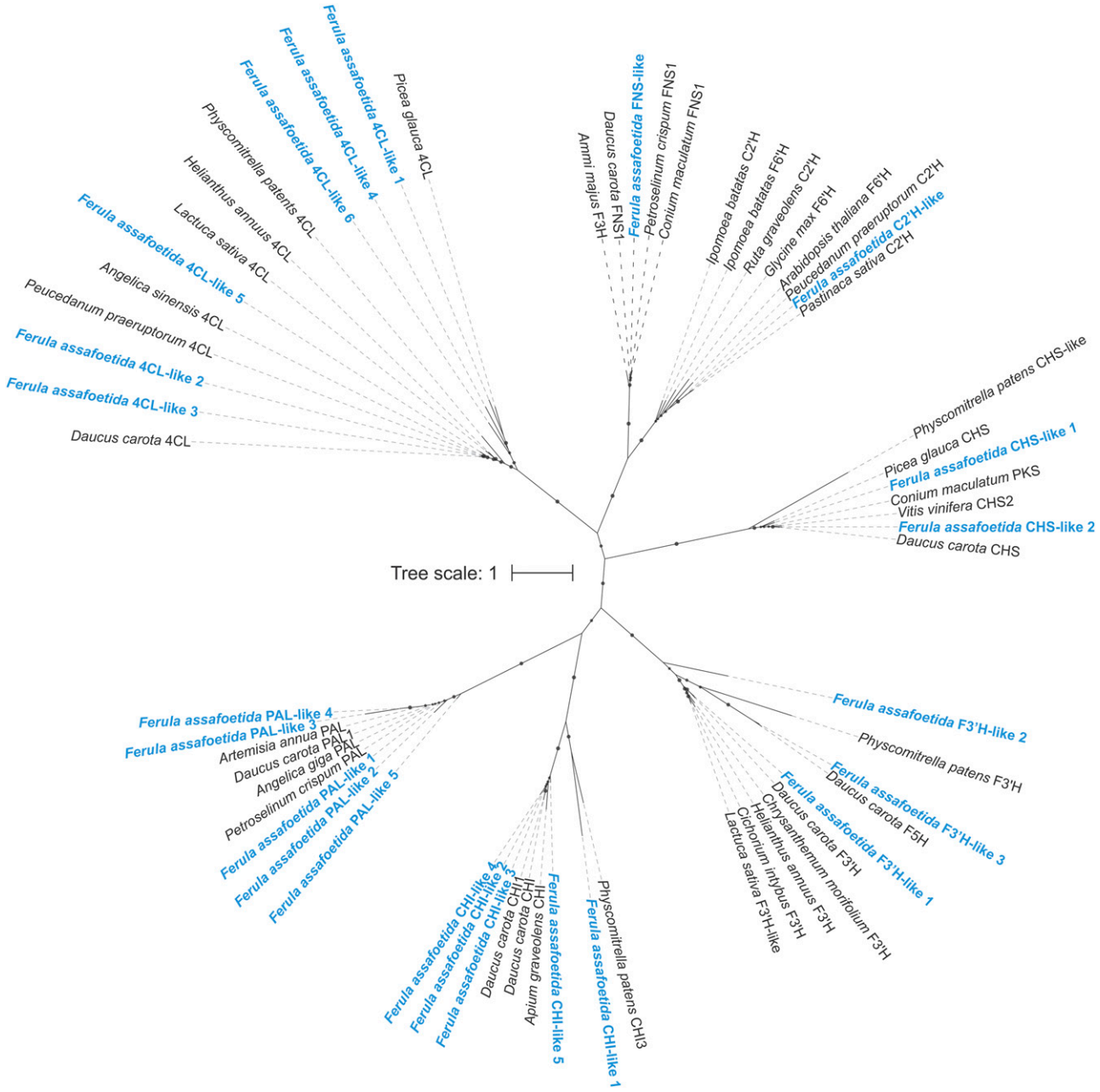

Figure 3 Maximum-likelihood phylogenetic tree illustrating interrelations relationships of enzyme candidates identified in *F. assafoetida* with known phenylpropanoid-biosynthetic enzymes. Branches with bootstrap support of >80% (500 repetitions) are highlighted with black dots. PAL, phenylalanine ammonia-lyase; 4CL, 4-coumarate:CoA-ligase; CHS, chalcone synthases; CHI, chalcone isomerases; FNS, flavone synthases; F3′H, flavanone 3-hydroxylase.

geraniol synthase DcTPS2, did not allow a more detailed functional prediction (Keilwagen *et al.* 2017).

Phylogenetic analysis of candidate genes for coumarin biosynthesis in *F. assafoetida* showed the presence of phenylalanine ammonia-lyase (PAL) and 4-coumarate:CoA-ligase (4CL) as key enzymes controlling phenylpropanoid biosynthesis as multi-gene families of five and six members, respectively (Figure 3). Although the size of the 4CL family in *F. assafoetida* is unclear with 4CL-like 4 and 6 being represented as partial sequences only, 4CL enzymes in other species are more commonly encoded by a single gene (Ehlting *et al.* 1999; Vogt 2010). Hence, a more expansive evolutionary divergence of this pathway component may have occurred in *F. assafoetida*. In contrast, only a single transcript was identified that showed significant similarity to p-coumaroyl-CoA 2′-hydroxylase (C2′H) enzymes that catalyze the hydroxylation of the 4CL product p-coumaroyl-CoA as a key reaction in the formation of coumarins such as umbelliferone (Yao *et al.* 2017) (Figure 3). Unlike coumarins, biosynthesis of flavone metabolites, including luteolin abundant in *F. assafoetida* oleo-gum-resin and other organs (Figure S7), proceeds through the activity of a chalcone synthase (CHS), followed by further modifications by chalcone isomerases (CHI), flavone synthases (FNS), and flavanone 3-hydroxylases (F3′H) (Bourgaud *et al.* 2006; Naoumkina *et al.* 2010; Vogt 2010). Phylogenetic analyses of the

identified phenylpropanoid-metabolic gene candidates in *F. assafoetida* showed small gene families of CHS, CHI and F3′H enzymes the occurrence of these pathway enzymes as multi-gene families in other plant species (Naoumkina *et al.* 2010; Vogt 2010). Conversely, FNS appears to be encoded by a single gene with the highest similarity to known type I FNS dioxygenases rather than FNSII enzymes of the CYP93B P450 subfamily in members of the Apiaceae (Britsch 1990; Fliegmann *et al.* 2010).

Tissue-specific differential expression analysis

The highest levels of differential gene expression were observed between leaves and roots, with a total of 3210 genes differentially expressed (FDR <0.05). Surprisingly, the flower *vs.* root contrast showed substantially fewer genes differentially expressed (420; FDR < 0.05) than leaf *vs.* root; this may be due to greater sample heterogeneity in the flower samples as compared to the leaf samples. The Venn diagram of pairwise comparisons indicated the highest similarly belonged to differential expressed genes of flowers compared to roots and leaves compared to roots (275) (Supplementary Figure S8).

Over-representation analysis of differentially expressed genes

The differentially expressed genes in different organs are indicative of biological function most relevant to each organ. Thus, we investigated the GO term over-representation of differentially expressed genes among select pairwise organ comparisons (Supplementary Figure S9). Photosynthesis GO terms were significantly upregulated in leaves (p-adjust $<1e^{-09}$) and flowers (p-adjust $<4e^{-05}$) as compared to roots (Figure S9A and S9B). With respect to terpenoid we observed terpene synthesis activity GO term as over-represented in up-regulated genes in roots *vs.* flowers (p-adjust <0.007) and up-regulated in flowers *vs.* stems (p-adjust <0.0002), suggesting that roots have the highest expression of the GO category, followed by flowers, and then stems (Figure S9C).

To examine the tissue-specificity of the biosynthetic pathways of the metabolites targeted here, we further studied the over-represented KEGG orthology (KO) terms in differentially expressed genes using GOseq package. The pathway analysis using KEGG database indicated several over-represented KO terms for the differential expressed genes in various organs (Supplementary Figure S10). We found that the KEGG pathway for biosynthesis of sesqui- and tri-terpenoid was upregulated in flowers *vs.* stems while the KO term for flavonoid biosynthesis was upregulated in flowers *vs.* other organs (Supplementary Figure S10). These two pathways were visualized, and the candidate genes were colored (Supplementary Figure S11 and S12).

Gene network analysis (WGCNA)

An alternative to identifying gene involved in terpenoid and flavonoid biosynthesis is to use gene co-expression to reconstruct genetic modules. Because this method treats each sample separately it may be better at identifying clusters of genes that function together. To accomplish this, we performed a weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath 2008) to find genetic modules which are highly co-expressed across the 12 samples (3 samples from 4 different organs) in *F. assafoetida*. WGCNA found 43 non-overlapping modules ranging from 38 to 3557 total gene size (Supplementary Figure S13).

Our metabolite analysis had indicated substantial variation in terpenoid and flavonoid abundance among different organs (Supplementary Figure S6 and S7). To find modules associated with these patterns, we next asked if any module could be related to our metabolites of interest. To do this, we calculated the correlation of each module's eigengene expression value and the measured metabolites of interest: terpenoid and coumarin-type phenylpropanoids including umbelliferone and luteolin.

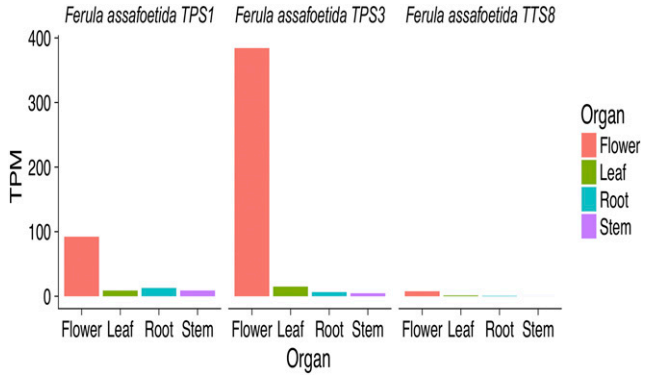

Figure 4 The TPM (Transcripts Per Kilobase Million values) of the candidate genes of *TPS*, terpene synthases; *TTS*, triterpenoid metabolism for different organs. The candidate gene names were provided in Supplementary Table S5.

One module of interest is the darkseagreen3 module. The darkseagreen3 module eigengene had significant correlations with sesquiterpenoid (r= -0.76 and adjusted p–value <0.0038) and the luteolin flavone compound (r= 0.83 and adjusted p–value <0.0014). Furthermore, the GO term significantly associated with the darkseagreen3 module was terpene synthase activity (Supplementary Figure S14A). Combined with the correlation with sesquiterpenoid content (adjusted p–value <0.0038), it is clear that genes in this module are important for sesquiterpenoid biosynthesis. The correlation coefficient (r= -0.76) indicated negative correlations between the darkseagreen3 module eigengene and sesquiterpenoid, meaning that higher expression of this module corresponds to less sesquiterpene synthesis. This suggests that the darkseagreen3 module contains genes which act as repressors for sesquiterpene synthesis or that shunt precursors into alternative pathways.

The darkseagreen3 module also was over-represented in GO terms "transferase activity, transferring hexosyl groups" that are parent of the "Luteolinidin 5-O-glucosyltransferase activity" term. This suggests that the darkseagreen3 module could be related to luteolin biosynthesis (Figure S14A). Since the flowers had the highest level of Luteolin while roots had the highest level of sesquiterpenoid, it could be concluded the darkseagreen3 module regulating the balance between luteolin and sesquiterpene biosynthesis.

Having identified the darkseagreen3 module as being associated with sesquiterpenoid and flavonoid biosynthesis we next investigated the differential expression patterns of the identified candidate genes for terpenoid and flavonoid metabolism. Note that some of these candidate genes had very low expression abundance and were not considered in differential expression analysis. Among the candidate genes of the terpenoid pathway, *F. TPS1*, *TPS3*, and *TTS8*, were located in darkseagreen3 module. These candidate genes of terpenoid pathway, were more abundant in flowers as compared to all other organs tested (Figure 4, Supplementary Figure S11 and Supplementary Table S7). This observation was consistent with the eigengene values for upregulated sesqui- and tri-terpenoid biosynthetic pathways among different organs (Figure S14B). Thus, these are candidate genes for the high levels of sesqui- and tri-terpenoid compounds observed in flowers.

Consistent with luteolin being most abundant in flowers (Supplementary Figure S7B) (Naoumkina *et al.* 2010; Vogt 2010; Li *et al.* 2018),

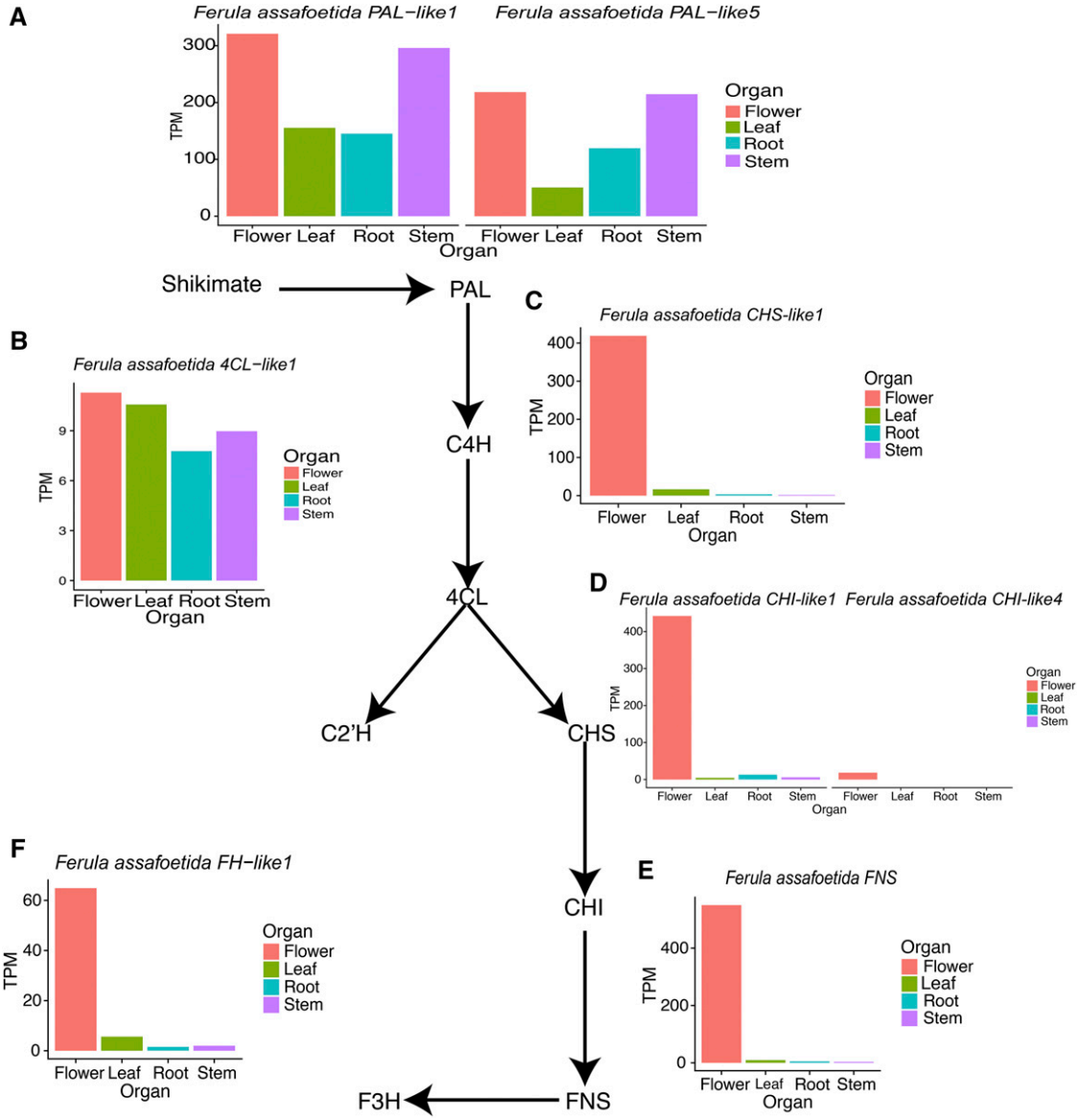

Figure 5 The TPM (Transcripts Per Kilobase Million values) of the candidate genes of *PAL* (A), *4CL* (B), *CHS* (C), *CHI* (D), *FNS* (E), *F3H* (F) as key genes of phenylalanine reactions for different organs. *PAL*, phenylalanine ammonia-lyase; *4CL*, 4-coumarate:CoA-ligase; *CHS*, chalcone synthases; *CHI*, chalcone isomerases; *FNS*, flavone synthases; and *F3′H*, flavanone 3-hydroxylase. The candidate gene names were provided in Supplementary Table S6.

we identified several candidate genes including *F. CHI-like1*, *CHI-like4*, *CHS-like1*, *FH-like1*, and *FNS* with predicted roles in phenylpropanoid metabolism that were most abundant in flowers and were also located in the darkseagreen3 WGCNA module (Figure 5). Note that, *F. assafoetida CHI-like4*, *FH-like1*, and *FNS* were upregulated in flavonoid biosynthesis pathway (Supplementary Figure S12). FNS activity is essential for luteolin biosynthesis (Chen *et al.* 2018), thus, *FNS* could be considered as candidate gene for luteolin biosynthesis in *F. assafoetida*. All these candidates were upregulated in flowers (Figure 5, Supplementary Figure S12 and Supplementary Table S8).

Also of interest, is the coral module. The coral module has a positive and significant correlation (r = 0.72, adjusted p-value = 0.008) with sesquiterpene. The coral module exhibited higher abundance in roots than other organ types. This means that the genes in coral module may act as activators for sesquiterpene synthesis. See Figure S15 for a depiction of how the darkseagreen3 and coral modules may act in sesquiterpene biosynthesis. The coral module included three transcripts, "oases6_k43_Locus_7389_Transcript_8_1", "oases6_k31_Locus_24182_Transcript_3_1", and "oases6_CL10402Contig1_1", that had significant matches to terpene synthase genes. These three genes also have a GO term of terpene synthase activity indicating their contribution to this function. Furthermore, these candidate genes all had a higher expression level in roots *vs.* flowers (Figure S16).

Combining tissue-specific transcriptome and metabolite analyses of the medicinal plant *F. assafoetida* identified candidate genes with possible roles in the biosynthesis of sesquiterpenoid and flavonoid metabolites as major bioactive constituents in the plants oleo-gum-resin. These resources can facilitate further gene function studies toward key bioactive natural products that define the medicinal properties of this traditional medicinal plant. Moreover, we provide detailed

assembly protocol to enable efficient transcriptome analyses in a broader range of non-model plant species.

ACKNOWLEDGMENTS

The authors acknowledge the Iran National Science Foundation (INSF, No. 95000275) for the financial support of part of this work. We especially thank Dr. C. Titus Brown, Kazunari Nozue, Ruijuan Li, and John Davis for helpful discussion.

LITERATURE CITED

Ahmed, A. A., M. M. Bishr, M. A. El-Shanawany, E. Z. Attia, S. A. Ross *et al.*, 2005 Rare trisubstituted sesquiterpenes daucanes from the wild *Daucus carota*. Phytochemistry 66: 1680–1684. https://doi.org/10.1016/j.phytochem.2005.05.010

Amalraj, A., and S. Gopi, 2017 Biological activities and medicinal properties of Asafoetida: A review. J. Tradit. Complement. Med. 7: 347–359. https://doi.org/10.1016/j.jtcme.2016.11.004

Anisimova, M., M. Gil, J. F. Dufayard, C. Dessimoz, and O. Gascuel, 2011 Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60: 685–699. https://doi.org/10.1093/sysbio/syr041

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate - a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B. Stat. Methodol. 57: 289–300. Available at::://A1995QE45300017.

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

Bourgaud, F., A. Hehn, R. Larbat, S. Doerper, E. Gontier *et al.*, 2006 Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. Phytochem. Rev. 5: 293–308. https://doi.org/10.1007/s11101-006-9040-2

Bray, N. L., H. Pimentel, P. Melsted, and L. Pachter, 2016 Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34: 525–527 (erratum: Nat. Biotechnol. 34: 888). https://doi.org/10.1038/nbt.3519

Britsch, L., 1990 Purification and characterization of flavone synthase I, a 2-oxoglutarate-dependent desaturase. Arch. Biochem. Biophys. 282: 152–160. https://doi.org/10.1016/0003-9861(90)90099-K

Cabau, C., F. Escudié, A. Djari, Y. Guiguen, J. Bobe *et al.*, 2017 Compacting and correcting Trinity and Oases RNA-Seq de novo assemblies. PeerJ 5: e2988. https://doi.org/10.7717/peerj.2988

Chen, F., D. Tholl, J. Bohlmann, and E. Pichersky, 2011 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. 66: 212–229. https://doi.org/10.1111/j.1365-313X.2011.04520.x

Chen, Z., G. Liu, N. Tang, and Z. Li, 2018 Transcriptome analysis reveals molecular signatures of luteoloside accumulation in senescing leaves of Lonicera macranthoides. Int. J. Mol. Sci. 19: 1012. https://doi.org/10.3390/ijms19041012

Conesa, A., S. Götz, J. M. García-Gómez, J. Terol, M. Talón *et al.*, 2005 Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Crusoe, M. R., H. F. Alameldin, S. Awad, E. Boucher, A. Caldwell *et al.*, 2015 The khmer software package: enabling efficient nucleotide sequence analysis. F1000 Res. 4: 900. https://doi.org/10.12688/f1000research.6924.1

Curini, M., G. Cravotto, F. Epifano, and G. Giannone, 2006 Chemistry and biological activity of natural and synthetic prenyloxycoumarins. Curr. Med. Chem. 13: 199–222. https://doi.org/10.2174/092986706775197890

Divya, K., K. Ramalakshmi, P. S. Murthy, and L. Jagan Mohan Rao, 2014 Volatile oils from *Ferula asafoetida* varieties and their antimicrobial activity. Lebensm. Wiss. Technol. 59: 774–779. https://doi.org/10.1016/j.lwt.2014.07.013

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21. https://doi.org/10.1093/bioinformatics/bts635

Ehlting, J., D. Büttner, Q. Wang, C. J. Douglas, I. E. Somssich *et al.*, 1999 Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. Plant J. 19: 9–20. https://doi.org/10.1046/j.1365-313X.1999.00491.x

Ewels, P., M. Magnusson, S. Lundin, and M. Käller, 2016 MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32: 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Falara, V., and E. Pichersky, 2012 Plant volatiles and other specialized metabolites: synthesis, storage, emission, and function, pp. 109–123 in *Secretions and Exudates in Biological Systems. Signaling and Communication in Plants*, edited by Vivanco, J., and F. Baluška. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23047-9_6

Falara, V., J. M. Alba, M. R. Kant, R. C. Schuurink, and E. Pichersky, 2014 Geranyllinalool synthases in solanaceae and other angiosperms constitute an ancient branch of diterpene synthases involved in the synthesis of defensive compounds. Plant Physiol. 166: 428–441. https://doi.org/10.1104/pp.114.243246

Finn, R. D., P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry *et al.*, 2015 The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44: D279–D285. https://doi.org/10.1093/nar/gkv1344

Fliegmann, J., K. Furtwängler, G. Malterer, C. Cantarello, G. Schüler *et al.*, 2010 Flavone synthase II (CYP93B16) from soybean (*Glycine max* L.). Phytochemistry 71: 508–514. https://doi.org/10.1016/j.phytochem.2010.01.007

Götz, S., J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj *et al.*, 2008 High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36: 3420–3435. https://doi.org/10.1093/nar/gkn176

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29: 644–652. https://doi.org/10.1038/nbt.1883

Habegger, R., and W. H. Schnitzler, 2000 Aroma compounds in the essential oil of carrots (*Daucus carota* L. ssp. sativus). 1. Leaves in comparison with roots. J. Appl. Bot. 74: 220–223.

Herde, M., K. Gärtner, T. G. Köllner, B. Fode, W. Boland *et al.*, 2008 Identification and regulation of TPS04/GES, an Arabidopsis geranyllinalool synthase catalyzing the first step in the formation of the insect-induced volatile C16-homoterpene TMTT. Plant Cell 20: 1152–1168. https://doi.org/10.1105/tpc.106.049478

Iorizzo, M., S. Ellison, D. Senalik, P. Zeng, P. Satapoomin *et al.*, 2016 A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat. Genet. 48: 657–666. https://doi.org/10.1038/ng.3565

Iranshahi, M., F. Kalategi, R. Rezaee, A. R. Shahverdi, C. Ito *et al.*, 2008 Cancer chemopreventive activity of terpenoid coumarins from *Ferula* species. Planta Med. 74: 147–150. https://doi.org/10.1055/s-2008-1034293

Iranshahy, M., and M. Iranshahi, 2011 Traditional uses, phytochemistry and pharmacology of asafoetida (*Ferula assa-foetida* oleo-gum-resin)-a review. J. Ethnopharmacol. 134: 1–10. https://doi.org/10.1016/j.jep.2010.11.067

Kalvari, I., J. Argasinska, N. Quinones-Olvera, E. P. Nawrocki, E. Rivas *et al.*, 2017 Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 46: D335–D342. https://doi.org/10.1093/nar/gkx1038

Kanehisa, M., Y. Sato, and K. Morishima, 2016 BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428: 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kavoosi, G., and V. Rowshan, 2013 Chemical composition, antioxidant and antimicrobial activities of essential oil obtained from *Ferula assa-foetida* oleo-gum-resin: Effect of collection time. Food Chem. 138: 2180–2187. https://doi.org/10.1016/j.foodchem.2012.11.131

Keilwagen, J., H. Lehnert, T. Berner, H. Budahn, T. Nothnagel *et al.*, 2017 The terpene synthase gene family of carrot (*Daucus carota* L.):

identification of qTLs and candidate genes associated with terpenoid volatile compounds. Front. Plant Sci. 8: 1930. https://doi.org/10.3389/fpls.2017.01930

Khan, A., and A. Mathelier, 2017 Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics 18: 287. https://doi.org/10.1186/s12859-017-1708-7

Kitaoka, N., X. Lu, B. Yang, and R. J. Peters, 2015 The application of synthetic biology to elucidation of plant mono-, sesqui- and di-terpenoid metabolism. Mol. Plant 8: 6–16. https://doi.org/10.1016/j.molp.2014.12.002

Langfelder, P., and S. Horvath, 2008 WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559. https://doi.org/10.1186/1471-2105-9-559

Li, Y., J. Fang, X. Qi, M. Lin, Y. Zhong *et al.*, 2018 Combined analysis of the fruit metabolome and transcriptome reveals candidate genes involved in flavonoid biosynthesis in *Actinidia arguta*. Int. J. Mol. Sci. 19: 1471. https://doi.org/10.3390/ijms19051471

Luo, W., and C. Brouwer, 2013 Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics 29: 1830–1831. https://doi.org/10.1093/bioinformatics/btt285

Luo, Y., P. Shang, and D. Li, 2017 Luteolin: A flavonoid that has multiple cardio-protective effects and its molecular mechanisms. Front. Pharmacol. 8: 692. https://doi.org/10.3389/fphar.2017.00692

Matsumoto, S., M. Mizutani, K. Sakata, and B. I. Shimizu, 2012 Molecular cloning and functional analysis of the ortho-hydroxylases of p-coumaroyl coenzyme A/feruloyl coenzyme A involved in formation of umbelliferone and scopoletin in sweet potato, *Ipomoea batatas* (L.) Lam. Phytochemistry 74: 49–57. https://doi.org/10.1016/j.phytochem.2011.11.009

Naoumkina, M. A., Q. Zhao, L. Gallego-Giraldo, X. Dai, P. X. Zhao *et al.*, 2010 Genome-wide analysis of phenylpropanoid defence pathways. Mol. Plant Pathol. 11: 829–846. https://doi.org/10.1111/j.1364–3703.2010.00648.x

Pangarova, T. T., and G. G. Zapesochnaya, 1973 Flavonoids of *Ferula assafoetida*. Chem. Nat. Compd. 9: 768. https://doi.org/10.1007/BF00565809

Pateraki, I., A. M. Heskes, and B. Hamberger, 2015 Cytochromes P450 for terpene functionalization and metabolic engineering, pp. 107–139 in: *Biotechnology of Isoprenoids. Advances in Biochemical Engineering/Biotechnology*, edited by Schrader, J., and J. Bohlmann. Springer, Cham. https://doi.org/10.1007/10_2014_301

Peters, R. J., 2010 Two rings in them all: The labdane-related diterpenoids. Nat. Prod. Rep. 27: 1521. https://doi.org/10.1039/c0np00019a

Pruitt, K. D., T. Tatusova, and D. R. Maglott, 2006 NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res., 35(suppl_1): D61–D65.

R Core Team, 2016 R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Richter, A., C. Schaff, Z. Zhang, A. E. Lipka, F. Tian *et al.*, 2016 Characterization of biosynthetic pathways for the production of the volatile homoterpenes DMNT and TMTT in *Zea mays*. Plant Cell 28: 2651–2665. https://doi.org/10.1105/tpc.15.00919

Robinson, M. D., and A. Oshlack, 2010a A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11: R25. https://doi.org/10.1186/gb-2010-11-3-r25

Robinson, M. D., D. J. McCarthy, and G. K. Smyth, 2010b edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140. https://doi.org/10.1093/bioinformatics/btp616

Schulz, M. H., D. R. Zerbino, M. Vingron, and E. Birney, 2012 Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28: 1086–1092. https://doi.org/10.1093/bioinformatics/bts094

Scott, C., 2016 Dammit: an open and accessible de novo transcriptome annotator. in prep. [Internet]; Available from: www.camillescott.org/dammit.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Suzek, B. E., Y. Wang, H. Huang, P. B. McGarvey, C. H. Wu *et al.*, 2014 UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31: 926–932. https://doi.org/10.1093/bioinformatics/btu739

Tenenbaum, D., 2018 KEGGREST: Client-side REST access to KEGG. R package version 1.20.0.

Tohge, T., M. Watanabe, R. Hoefgen, and A. R. Fernie, 2013 The evolution of phenylpropanoid metabolism in the green lineage. Crit. Rev. Biochem. Mol. Biol. 48: 123–152. https://doi.org/10.3109/10409238.2012.758083

Townsley, B. T., M. F. Covington, Y. Ichihashi, K. Zumstein, and N. R. Sinha, 2015 BrAD-seq: Breath Adapter Directional sequencing: a streamlined, ultra-simple and fast library preparation protocol for strand specific mRNA library construction. Front. Plant Sci. 6: 366. https://doi.org/10.3389/fpls.2015.00366

Vialart, G., A. Hehn, A. Olry, K. Ito, C. Krieger *et al.*, 2012 A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70: 460–470. https://doi.org/10.1111/j.1365-313X.2011.04879.x

Vogt, T., 2010 Phenylpropanoid biosynthesis. Mol. Plant 3: 2–20. https://doi.org/10.1093/mp/ssp106

Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis *et al.*, 2018 BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. 35: 543–548. https://doi.org/10.1093/molbev/msx319

Wurtzel, E. T., and T. M. Kutchan, 2016 Plant metabolism, the diverse chemistry set of the future. Science 353: 1232–1236. https://doi.org/10.1126/science.aad2062

Xiao, M., Y. Zhang, X. Chen, E. J. Lee, C. J. S. Barber *et al.*, 2013 Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. J. Biotechnol. 166: 122–134. https://doi.org/10.1016/j.jbiotec.2013.04.004

Yahyaa, M., D. Tholl, G. Cormier, R. Jensen, P. W. Simon *et al.*, 2015 Identification and Characterization of Terpene Synthases Potentially Involved in the Formation of Volatile Terpenes in Carrot (*Daucus carota* L.). Roots. J. Agric. Food Chem. 63: 4870–4878. https://doi.org/10.1021/acs.jafc.5b00546

Yahyaa, M., M. Ibdah, S. Marzouk, and M. Ibdah, 2018 Profiling of the terpene metabolome in carrot fruits of wild (*Daucus carota* L. ssp. carota) accessions and characterization of a geraniol synthase. J. Agric. Food Chem. 66: 2378–2386. https://doi.org/10.1021/acs.jafc.6b03596

Yao, R., Y. Zhao, T. Liu, C. Huang, S. Xu *et al.*, 2017 Identification and functional characterization of a p-coumaroyl CoA 2′-hydroxylase involved in the biosynthesis of coumarin skeleton from *Peucedanum praeruptorum* Dunn. Plant Mol. Biol. 95: 199–213. https://doi.org/10.1007/s11103-017-0650-4

Ye, J., L. Fang, H. Zheng, Y. Zhang, J. Chen *et al.*, 2006 WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 34: W293–W297. https://doi.org/10.1093/nar/gkl031

Young, M. D., M. J. Wakefield, G. K. Smyth, and A. Oshlack, 2010 Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 11: R14. https://doi.org/10.1186/gb-2010-11-2-r14

Zdobnov, E. M., F. Tegenfeldt, D. Kuznetsov, R. M. Waterhouse, F. A. Simão *et al.*, 2016 OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. 45: D744–D749. https://doi.org/10.1093/nar/gkw1119

Zerbe, P., B. Hamberger, M. M. S. Yuen, A. Chiang, H. K. Sandhu *et al.*, 2013 Gene discovery of modular diterpene metabolism in nonmodel systems. Plant Physiol. 162: 1073–1091. https://doi.org/10.1104/pp.113.218347

Zerbe, P., and J. Bohlmann, 2015 Plant diterpene synthases: exploring modularity and metabolic diversity for bioengineering. Trends Biotechnol. 33: 419–428. https://doi.org/10.1016/j.tibtech.2015.04.006

Zerbino, D. R., and E. Birney, 2008 Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18: 821–829. https://doi.org/10.1101/gr.074492.107

Communicating Editor: J. Udall