J. Agric. Sci. Technol. (2024) Vol. 26(1): 177-192

Terpenoid Biosynthetic Pathway in *Ferula persica* Using Transcriptome Analysis and Metabolome Data

J. Nasiri[1], A. Soorni[2], A. D. J. van Dijk[3], and M. R. Naghavi[4]*

ABSTRACT

An effort was made to analyze metabolome and transcriptome profiles of *Ferula persica* via GC-MS and RNA-seq data. The analysis of the essential oils extracted from both flower and root tissues demonstrated the prominence of monoterpene constituents, while sesquiterpene compounds were present in the lower magnitudes. Considering transcriptome analysis, 2127 differentially expressed genes were found between root and flower: 396 transcripts were up-regulated in root, while 1731 exhibited an up-regulation pattern in flower. Out of 2127 transcripts, 86 were annotated as Terpene Synthases (TPSs), of which 83 TPSs were classified subsequently into five individual sub-families of TPS-a (33), TPS-b (42), TPS-c (2), TPSe-f (3), and TPS-g (3). Several transcription factor families were recognized among the differentially expressed genes, suggesting their direct or indirect regulatory roles for the biosynthesis of terpenoids in *F. persica*. Finally, according to our phylogenetic results, both *F. assa-foetida* and *F. gummosa* were placed in the same clade, while *F. persica* was lonely settled in one monophyletic clade, with the estimated divergence time of 2.99 Million Years Ago (MYA) between *F. gummosa* and *F. assa-foetida*, and 3.87 MYA between *F. persica* and two other *Ferula* species.

Keywords: Ferula persica,Genome evolution, Medicinal plant, Phylogenetic results, RNA-Seq.

INTRODUCTION

Among various medicinal plants, the genus *Ferula* has been utilized as an herbaceous perennial plant species. The genus *Ferula* comprises ~172 species, distributed geographically from central Asia westward throughout the Mediterranean region to northern Africa (Kavoosi and Rowshan, 2013). In Iran, ~30 species of *Ferula* spp. have been recorded, some of which, including *F. tabasensis*, *F. gummosa*, and *F. persica,* are argued to be endemic of Iran, followed by *F. assa-foetida* which grows as a native plant in Kashmir, Iran and Afghanistan (Asili *et al*., 2009). The genera of *F. gummosa*, *F. persica* and *F. assa-foetida* are generally able to generate an invaluable mixture known as "*oleo-gum–resin*", which is commonly extracted from the exudates of the rhizome or taproot of the plants (Kavoosi and Rowshan, 2013). In the traditional and official markets, the following two types of oleo-gum-resin of asafoetida and sagapenum are available, collected normally from *F. assa-foetida*, *F. persica*, *F. foetida* and *F. alliacea*. Iran, followed by India and Afghanistan, is assumed as a major producer of Asafoetida, as its total production value has been estimated as 172,590 kg in 2018, with the market value of 3,701,447 $US (Barzegar *et al*., 2020).

Despite diverse investigations focusing on

[1] Nuclear Agriculture Research School, Nuclear Science and Technology Research Institute, AEOI, P. O. Box: 31485-498, Karaj, Islamic Republic of Iran.
[2] Department of Biotechnology, College of Agriculture, Isfahan University of Technology, Isfahan 84156-83111, Islamic Republic of Iran.
[3] Plant Sciences Group, Wageningen University, Wageningen, The Netherlands.
[4] Division of Biotechnology, Department of Agronomy and Plant Breeding, Agricultural and Natural Resources College, University of Tehran, P. O. Box: 31587-11167, Karaj, Islamic Republic of Iran.
*Corresponding author; e-mail: mnaghavi@ut.ac.ir

the simultaneous analyses of both transcriptome and metabolome data in other plant species (Barzegar *et al*., 2020), followed by various studies related to the chemical composition, pharmacological effect, and green chemistry (Nasiri *et al*., 2019; Nasiri *et al*., 2018), little genomic and transcriptomic information are available for the genus *Ferula,* including *F. gummosa* (Najafabadi *et al*., 2017) and *F. assa-foetida* (Amini *et al*., 2019). However, no relevant investigation has been reported on the subject of *F. persica*. Therefore, an endeavor was made to analyze transcriptome and metabolome data set of *F. persica*, aiming to acquire more insights about different regulatory mechanisms governing the plant concerning various metabolic pathways.

MATERIALS AND METHODS

Plant Materials

Three individual plant samples of *F. persica* were collected in May 24, 2017 from the mountainous region of Mowrud Village, Pol-e Khab with an altitude of 2,365 M Above Sea Level (MASL; 36° 00′ 06″ N and 51° 12′ 06″ E), Karaj, Alborz Province, Iran. The formal identification of the plant material was undertaken by the Herbarium of Agricultural and Natural Resources College, University of Tehran. Two parts of the harvested plants including roots and flowers were gathered and frozen immediately in liquid nitrogen and kept at -80°C.

Preparation of Plant Extracts

Upon air dying, for each tissue, three replicates were pooled, and ~20 g of the fine powder were weighed and utilized for essential oil isolation through hydro distillation for 5 hours, using a Clevenger type apparatus. The resultant essential oils of both tissues were kept at 4°C until injected into GC-MS (Amini *et al*., 2019)..

GC-MS Analysis

GC-MS analysis was performed via an Agilent GC, equipped with mass selective detector with quadrupole analyzer MD800. The electron ionization energy was 70eV, ion-source at 200°C and the interface temperature of 280°C. A split-split less injection (split ratio 1: 10) at 280°C injector temperature was employed**.** A fused silica column 5% phenyl-poly-dimethyl-siloxane (Chrompak CP-Sil 8 CB 50 m×250 µm×0.12 µm**)** was utilized. The oven temperature was programmed as follows: from 50°C (2 minutes hold) raised at 4°C min^{-1} to 120°C, then, raised at 2°C min^{-1} to 200°C, then, raised at 25°C min^{-1} to 280°C (8 minutes hold). A sample of 1.0 µL was injected. Data acquisition was performed with Mass Lab software for the mass ranges 30-600 u with a scan speed of 1.0 scan/second**.** The identification of compounds was also based on the Kovats retention indices. The components were identified by comparison of their mass spectra with data from Adams, US National Institute of Standards and Technology (NIST, USA), WILEY 1996 Ed. Mass Spectra Library (Amini *et al*., 2019).

RNA Extraction

One hundred mg of both tissues (two biological replications) were first homogenized in a mortar with liquid nitrogen, and, subsequently, RNA isolation was carried out using a pBIOZOL reagent (Invitrogen) according to manufacturer's instructions. Putative genomic DNA contamination was removed through treatment with *DNase* I (RNase- Free DNase Set, Fermentase). The quality and quantity of isolated RNAs were determined first using a NanoDrop® ND-1000 spectrophotometer. Furthermore, RNA Integrity (RIN) of the samples were validated at Macrogen Company (Korea), and those with RIN more than 8.0 were selected for RNA sequencing. The cDNA

libraries were constructed according to the instructions given in TruSeq Stranded mRNA LT Sample Prep Kit and then were subjected to sequencing on an Illumina HiSeq 2000 paired-end 151bp system.

Data Filtering and *de novo* Assembly

Following sequencing of libraries, the quality of reads was evaluated by using FastQC (v.0.11.8; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), then, adapter fragments and poor quality reads were removed using Trimmomatic v0.30 (Bolger *et al*., 2014). Parameters included Illumina clip with seed mismatches 2, palindrome clip threshold 30, simple clip threshold 10, leading quality and trailing quality 3, sliding window trimming with a window size 4, required quality 20, and minimum read length of 50 bp. Subsequently, *de novo* transcriptome assembly of clean reads was performed using the Trinity Program (Grabherr *et al*., 2011). The transcriptome assembly was further subject to the EvidentialGene tr2aacds pipeline (http://eugenes.org/EvidentialGene/) to remove redundant transcripts and obtain an 'optimal' set of *de novo* assembled transcripts. Finally, assembly was assessed using the script TrinityStats.pl contained in Trinity package and BUSCO v.3 (Benchmarking Universal Single-Copy Orthologs) (Simão *et al*., 2015) to acquire the percentage of single-copy orthologs represented in Eukaryota dataset.

Expression Profile Analysis and TPSs Identification

The RNA-Seq by Expectation Maximization (RSEM) method (Li and Dewey, 2011) with the default parameters was used to quantify gene expression level. Clean reads of each library were mapped back onto the assembled transcriptome, then, read count from all samples were combined into a matrix using script abundance_estimates_to_matrix.pl. After producing reads count matrix, Pearson's correlation coefficient between each pair of biological replicates was evaluated by comparing $\log_{10}$ of FPKM values. Finally, the Differentially Expressed Genes (DEGs) were analyzed through the IDEAMEX website (Jimenez-Jacinto *et al*., 2019), using the DESeq2 (Love *et al*., 2014), EdgeR (Robinson *et al*., 2010), NOISeq (Tarazona *et al*., 2011) and limma-Voom (Ritchie *et al*., 2015). The threshold to judge the signific`ance of gene expression differences was False Discovery Rate ''FDR≤ 0.01, Counts per million CPM= 3 and the absolute value of Log fold change logFC>= 2". Furthermore, TPS unigenes were identified from the transcriptome assembly using sequence homology (Priya *et al*., 2018). To compare TPS genes identified in *F. persica* against other *Ferula* species, transcriptome data for *F. gummosa* (BioPoject: PRJNA328267) and *F. asafoetida* (BioPoject: PRJNA476150) were obtained from the NCBI database and investigated to discover TPS genes. In addition, regular conserved motifs of TPSs were obtained using MEME tools.

TFs Identification and GO Classification of DEGs

Transcription Factors (TFs) were identified and classified using iTAK (http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi). The Gene Group Enrichment Analysis (GSEA) was performed using the GO terms in the agriGO v2.0 (Tian *et al*., 2017) software. TransDecoder and Trinotate software suites were used for functional annotation of each assembly following the method outlined at (http://trinotate.github.io/). For GO and gene set enrichment analysis, transcripts were mapped to the protein sequences source of Arabidopsis (Araport11_genes.201606.pep.fasta) using the BLAST search. This is because of the

well-maintained and annotated Arabidopsis genome. Finally, the Gene Group Enrichment Analysis (GSEA) was performed using the GO terms in the agriGO v2.0 software.

Single-Copy Orthogroups Identification for Comparative Phylogenetic Analysis

To find orthogroups among all the predicted protein sequences acquired from GeneMarkS-T v2.0.1, the program called OrthoFinder was employed (Emms and Kelly, 2015) in terms of default parameters. Meanwhile, in precisely rooting the resultant phylogenetic tree, the sequence assembly data of *Thapsia garganica* (SRP008179) and *Daucus carota* (Iorizzo *et al*., 2016) were utilized as outgroups. The orthogroups possessing just single copy genes were maintained for further analysis. The nucleotide sequences for each group were multiply aligned using MUSCLE v3.8.31. Subsequently, poorly aligned regions were filtered out via the trimAl v1.4 (Capella-Gutiérrez *et al*., 2009) on the basis of the parameter "-gt 0.9 -st 0.001". Then, a Bayesian model through BEAST v.2.5.2 was employed according to previous research (Soorni *et al*., 2019).

Primer Design and Quantitative Real Time PCR (qPCR)

To validate the expression pattern of genes involved in the terpenes biosynthetic pathway, qRT-PCR was applied. Gene-specific primers (Table S1) were designed using the IDTdna tools (http://www.idtdna.com), and a real-time PCR system (ABI ViiA 7 Real-time PCR) was employed, in a total reaction volume of 15 µL containing 7.5 µL SYBR Green Master Mix (BioFACT, Korea), 2 µL of diluted cDNA, and 1 µL of each primer (10 µM) in conjunction with adding PCR-grade water. The qPCR was carried out based on a thermal program of 5 min at 95°C, 40 cycles of 10 seconds at 95°C, 20 seconds at the specific annealing temperature for each primer, 20 seconds at 72°C, and, finally, a melting curve program. The *Actin* was used as an internal reference (housekeeping) gene. The statistical analysis of gene expression was conducted using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

RESULTS AND DISCUSSION

GC-MS Results

According to the GC-MS results conducted in the Iranian Institute of Medicinal Plants (IMP, Karaj, Iran), the essential oils of flower and root possessed 120 and 110 different bioactive compounds, respectively (Table S2-3). The most popular metabolites in flowers included monoterpene hydrocarbons (i.e., α-pinene, 6.68%, camphene, 3.24%; limonene, 1.39%), oxygenated monoterpenes (i.e., borneol, 2.26%, fenchyl acetate, 3.54%, bornyl acetate, 8.89%; L-Fenchone, 1.11%), 2,6-Octadien-1-ol, 3,7-dimethyl-, propanoate, (Z)- (10.06%), followed by butanoic acid, 3,7-dimethyl-2,6-octadienyl ester, (E)- (4.55%), 4-Mercaptoimidazo[4,5-c]pyridine (6.22%), and benzenecarbodithioic acid, 4-met hoxy-, ethyl ester (7.51%). (Table S2). In roots, however, monoterpene hydrocarbons (i.e., α-pinene, 25.76%, camphene, 11.96%, β-pinene, 1.89%, and D-Limonene, 4.56%), oxygenated monoterpenes (i.e., fenchyl acetate, 9.93%, isobornyl acetate, 3.37%), oxygenated Sesquiterpenes (i.e., carotol, 3.39%; beta-cedrene, 4.26%; and Di-epi-.alpha.-cedrene, 1.34%), and sesquiterpene hydrocarbons (i.e., sesquiphellandrene, 3.02%, and alpha.-farnesene, 1.43%) were the most represented metabolites.(Table S3).

Overall, our GC-MS analysis indicated that the essential oils of both flower and root tissues of *F. persica* were dominated by the monoterpenes fraction, while sesquiterpenes possessed lower quantities. These results were in contrast with a previous study on the

chemical composition analysis of *F. persica*, indicating the superiority of phenylpropanoides (64.7%) and sulfur compounds (28.6%) in aerial parts (Javidnia *et al*., 2005) and roots (Iranshahi *et al*., 2006), respectively. Nonetheless, in both aforementioned investigations, among different classes of terpenes, oxygenated monoterpenes were still the superior (13.0 and 23.2%, respectively).

RNA-seq and *de novo* Assembly

RNA-Seq of four libraries from flower and root tissues resulted in 119.86 million reads with more than 96 and 90% exhibiting quality score of Q20 and Q30, respectively. After trimming and removing poor and short reads, 90.02 million reads remained for the assembly (Table 1).

Using Trinity, clean reads were assembled into 204,433 transcripts, with a total length of 192.406 Mbp. The N50 value and mean length of these transcripts were 1434 and 941 bp respectively (Table 2). Subsequently, the tr2aacds pipeline was applied; comparing the resulting EvidentialGene set with the Trinity assembly indicated that the tr2aacds pipeline reduced the transcript number by 2.5 fold. The results of the percentage of reads mapping back to the final assembly ranged from 89.71 to 91.84%. The BUSCO values indicated that tr2aacds pipeline reduced the number of fragmented BUSCOs, while it increased the proportion of complete and single-copy BUSCOs. These results demonstrated that tr2aacds pipeline was capable of generating higher-quality transcripts by removing redundant or combining the high-quality transcripts.

Differentially Expressed Genes

The correlation results of replicates indicated that the estimated levels of gene expression were highly consistent between any replicate pair of each tissue (r= 0.91−0.92; Figure 1). Furthermore, a Multidimensional Scaling Plot (MDS; Figure S1) was designed based on the log2 Fold Change (logFC) expression to indicate the pattern of proximities among a set of *F. persica* objects. The tight clustering of the flower data points means there were fewer variations among biological replicates in comparison to the root samples. The MDS plot revealed two distinct clusters of flower and root samples, indicating high variability between the different tissues.

Differentially Expressed Genes (DEGs) were subsequently obtained between roots and flowers. In total, 6236, 6297, and 6940 DEGs

Table 1. Summary of RNA-seq data base quality from four RNA libraries of *F. persica.*

Original data				
Sample Name	Read Count	GC (%)	Q20[a] (%)	Q30[a] (%)
FlowerRep1	29,725,000	43.74	96.86	91.4
FlowerRep2	32,934,676	44.36	96.57	90.8
RootRep1	29,733,600	44.03	96.69	91.07
RootRep2	27,467,114	43.92	96.42	90.58
Clean data				
Sample Name	Read Count	GC (%)	Q20 (%)	Q30 (%)
FlowerRep1	22,597,182	43	98.08	93.51
FlowerRep2	22,626,130	43	97.85	92.75
RootRep1	22,506,218	43	98.05	93.65
RootRep2	22,295,040	43	98.01	93.43

[a] Q20 and Q30: The percentage of bases with a Phred value> 20 and > 30, respectively.

were obtained with the DESeq2, EdgeR and NOISeq methods, respectively, while only 2130 DEGs were obtained through applying limma method. Based on the results of Venn diagram (Figure S2), 2127 DEGs were shared among four methods. We selected DEGs that were validated as differentially expressed by all four methods for the subsequent analysis, since the results obtained by limma were almost entirely a subset of the result of each of the other methods. Among 2127 DEGs, 396 transcripts were up-regulated in root, while the remaining 1731 transcripts were up-regulated in flower.

Identification of *TPS* Genes and Conserved Motifs

Based on TERZYME analysis, 86 transcripts were nominated as TPSs among the entire assembly transcripts. According to the function-based classification, 41, 8, and 38 transcripts were defined as monoterpenes, diterpenes, and sesquiterpenes, respectively. Of which, only 11, one and three transcripts exhibited differential expression between root and flower samples, respectively (Figure 2).

To classify all the previously mentioned 86 TPSs into seven individual classes (from TPS-a to TPS-g) in terms of sequence homology-based method (TERZYME), a grouping analysis was also conducted. As the results indicated, 83 out of 86 TPSs (97%) were detached into five individual categories, while the remaining three TPSs failed to settle in a given class. The first group known as "TPS-a" contained 33 transcripts belonging to sesquiterpene family. Notably, three sesquiterpene genes identified as DEGs (see above) were included in the class of TPS-a. The second group known as "TPS-b"

Table 2. Assembly statistics results and BUSCO completeness assessment of *F. persica.*

	Trinity		tr2aacds	
	Transcript	Gene	Transcript	Gene
Total	204,433	98,297	82,302	56,010
Average contig length	941.17	764.69	978.27	994.56
Total assembled bases	192,406,550	75,166,392	80,513,557	55,705,063
Contig N50	1434	1261	1353	1450
Complete BUSCOs	90.5%		91.1%	
Complete and single-copy BUSCOs	35.0%		60.1%	
Complete and duplicated BUSCOs	55.5%		31.0%	
Fragmented BUSCOs	7.1%		3.6%	
Missing BUSCOs	2.4%		6.3%	

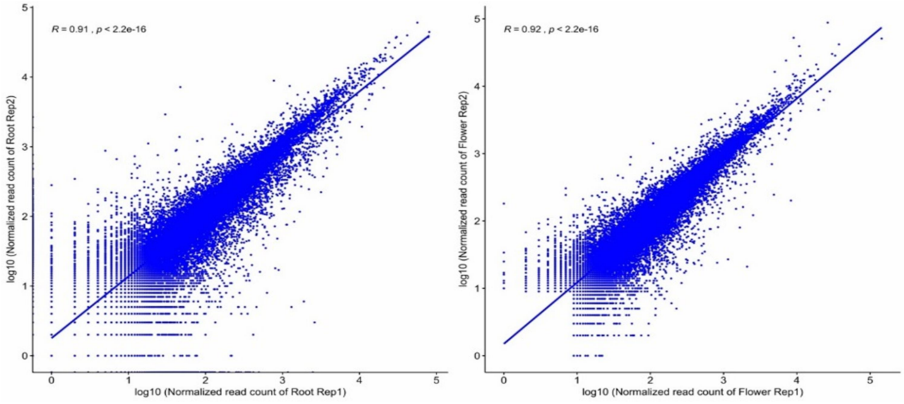

Figure 1. Plot of correlation between normalized read counts of biological replicates. Left: Correlation plot between root replicates, and Right: Correlation plot between flower replicates.

consisted of 42 transcripts belonging to mono (39 transcripts) and sesquiterpene (4 transcripts) families, which contained 11 differentially expressed monoterpenes. The third and fourth groups coined as "TPS-c" and "TPSe-f" encompassed two and three transcripts of diterpene family, respectively, without any responsibility in making differential expression pattern. Lastly, the group called "TPS-g" was nominated as the fifth class, containing two and one transcripts of mono and sesquiterpene families, respectively. As anticipated, most putative TPSs identified in the *Ferula* transcriptome were assigned to SES-TPS-a and Mono-TPS-b subfamily, supporting the existing view that the essential oils in *Ferula* species are dominated by the monoterpenes and sesquiterpenes fraction.

To compare our TPS results with the previous studies on the other *Ferula* species, the transcriptome assembly of both *F. gummosa* and *F. asafoetida* were also analyzed via TERZYME (Figure 3). In terms of the function-based classification, 52 TPSs were overall identified for *F. gummosa*, and, subsequently, classified as 15, 4, and 23 mono-, di-, and sesquiterpenes, respectively. Meanwhile, according to sequence homology-based method, 42 out of 52 TPSs were classified into four distinct sub-families *viz.* TPS-a (20), TPS-b (18), TPS-c (2), and TPSe-f (2). Considering *F. asafoetida*, 45 TPSs were overall identified, among which 16, 21, and 8 served as mono-, di-, and sesquiterpenes, respectively. Furthermore, taking sequence homology-based approach into account, all the 45 TPSs identified for *F. asafoetida* were categorized into 6 sub-families of TPS-a (4), TPS-b (16), TPS-c (16), TPS-d (5), TPSe-f (2), and lastly TPS-g (2).

It has been claimed that TPS-a, TPS-b and TPS-g subfamilies are angiosperm-specific, and among all, seven sub-families both b, and g followed by c gene subfamilies are contributed in secondary metabolism and

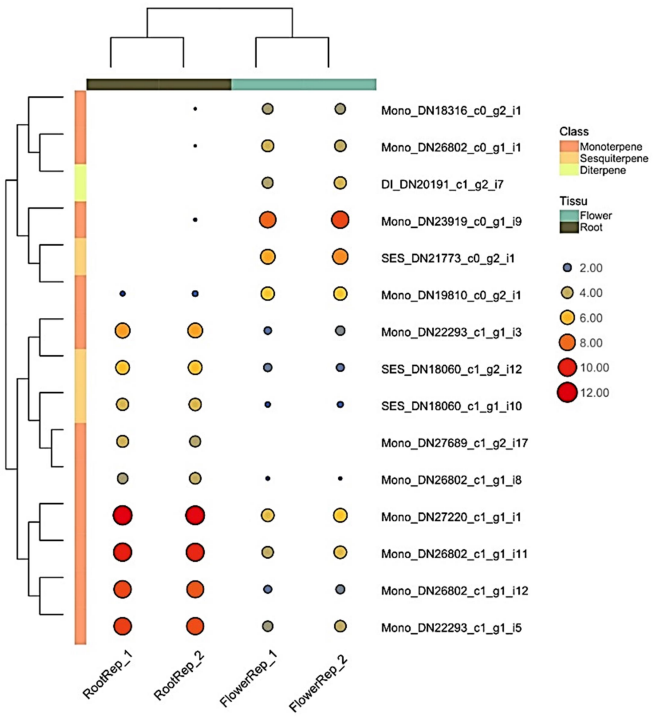

Figure 2. Heatmap of the expression levels and relationships of TPS genes across the two tissues of the *F. persica* transcriptome data. The color scale represents FPKM counts, and the ratios are log2 transformed.

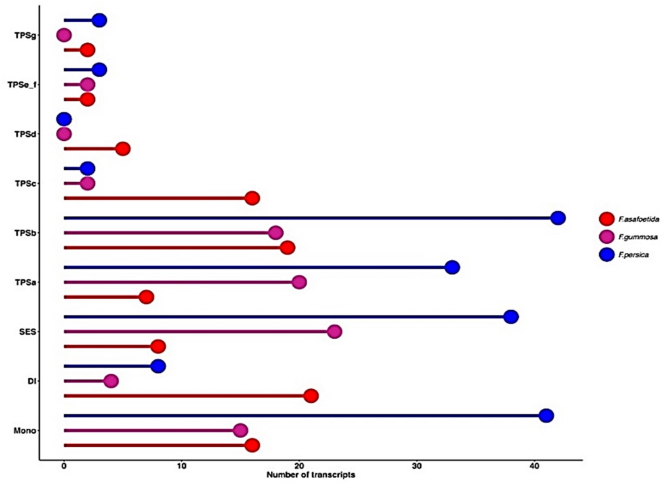

Figure 3. Distribution of TPS classes derived from three *Ferula* transcriptomes**.**

exhibited larger diversification. Interestingly, among various sub-families of TPSs in three species of *Ferula*, both TPS-b and TPS-g sub-families encompassed the highest frequencies as compared to the others, suggesting a high level of TPS-a- and TPS-b-mediated secondary metabolism processes, mainly terpenes in the genus *Ferula.*

On the other hand, the most TPS classes were identified in all the aforementioned three species of *Ferula*, suggesting their possible contributions in the generation of the structural and functional diversity of mono-, sesqui-, and diterpenes. Comparing GC-MS results to the TPS outputs, in *F. persica*, a good relationship was observed between the superiority of monoterpene quantities and monoterpene TPSs, indicating their key roles in producing monoterpenes in this species. Notably, the lowest TPSs counted for diterpenoid TPSs, and according to our current and earlier works, no diterpenes have yet been detected using metabolome profiling in the plant.

Identification of TFs

Among all the 2127 DEGs, 171 TFs belonging to 39 families, 19 Transcriptional Regulators, and 61 lastly Protein Kinases were overall identified. Among 39 families of TFs, the MYB family with the frequency value of 20 was the most popular one. The frequency of 10-top TF families is shown in Figure 4. To determine various TF families contributing to the biosynthesis regulation of secondary metabolites, several strategies have been recorded, among which RNA-Seq has been employed in countless studies worldwide. Notably, the majority of them have aimed to determine the type and frequency of TFs family, while the second investigation group have focused on characterizing those TFs regulating secondary metabolism, as well as their application in metabolic engineering of alkaloids and terpenoids (Wang and Gribskov, 2017; Yamada and Sato, 2013). Terpenes and terpenoids (terpene-like constituents) are one of the most diverse and largest known groups of PSMs (Ashour *et al*., 2018; Singh and Sharma, 2015), and represent one of the most important naturally occurring compounds in the genus *Ferula* spp. Therefore, due to their extensive distribution, it could be hypothesized that production/accumulation of terpenes is possibly regulated by various kinds of TF families. Similar results have been recorded in different plant species including both bHLH and MYB TFs in *A. thaliana* (Hong *et al*., 2012; Zvi *et al*., 2012), *Solanum lycopersicum* (Ji *et al*., 2014), *Artemisia annua* (Spyropoulou *et al*., 2014), followed by the WRKY family including GaWRKY1

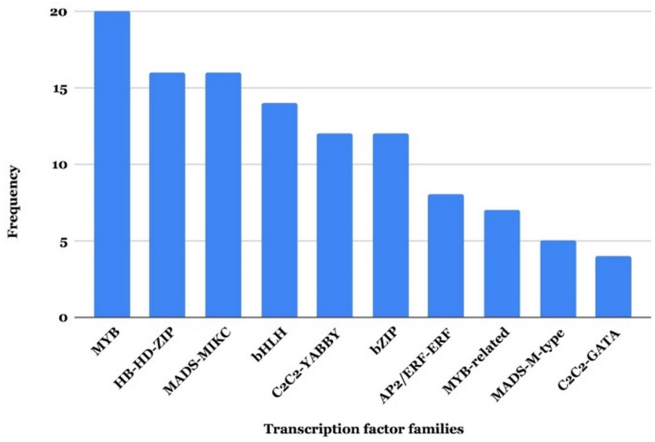

Figure 4. Distribution of DEGs in different TF families.

in *Gossypium arboreum* (Xu *et al*., 2004), SlMYC1 and SlWRKY73 in *S. lycopersicum* (Spyropoulou *et al*., 2014), and OsWRKY76 in *Oryza sativa* (Yokotani *et al*., 2013) for regulation of terpenoids production/accumulation, TcMYC2a in regulating taxol biosynthesis in *Taxus chinensis* (Zhang *et al*., 2018), and R2R3-MYB to regulate fragrance biosynthesis in lilies (*Lilium* spp.) (Yoshida *et al*., 2018). In this context, based on earlier investigations as well as our current results, the most popular TF families were identified as "MYB, HB-HD-ZIP, MADS-MIKC, bHLH, C2C2-YABBY, bZIP, AP2/ERF-ERF, MYB-related, MADS-M-type, and C2C2-GATA". It could be concluded that some of them (if not all) may directly/indirectly regulate the terpenoids biosynthesis in *F. persica*.

Gene Ontology (GO) and Gene Set Enrichment Analysis

The GO enrichment analysis of the 2127 DEGs identified 247 significantly (FDR< 0.05) enriched GO terms for the biological process, cellular component, and molecular function categories. In terms of biological processes, these DEGs were classed into 138 classifications. Within the biological process category, the enriched DEGs were mainly associated with the metabolic process (GO:0008152), cellular process (GO:0009987) and primary metabolic process (GO:0044238). Secondary metabolite process ranked as 47. According to molecular function, DEGs were divided into 50 classifications: the most represented molecular functions were the catalytic activity (GO:0003824), binding (GO:0005488) and hydrolase activity (GO:0016787). In cellular component category, DEGs were clustered into 59 classifications. The largest subcategories of the cellular components were intracellular (GO:0005622), cell (GO:0005623), and cell part (GO:0044464). We further analyzed the DEGs involved in the enriched biological process GO terms. For "metabolic process (GO:0008152)", most of genes encoded NAD(P)-binding Rossmann-fold (10 genes) and UDP-Glycosyltransferase (9 genes).

Detecting WGDs Using Pairwise *Ks* and Phylogenetic Analysis

The distribution of *Ks* values between pairs of *Ferula* paralogs reflected evidence of a possible WGD at *Ks*= 0.4 (Figure 5). Although our simulations indicated possible WGD in *Ferula* species, it was difficult to

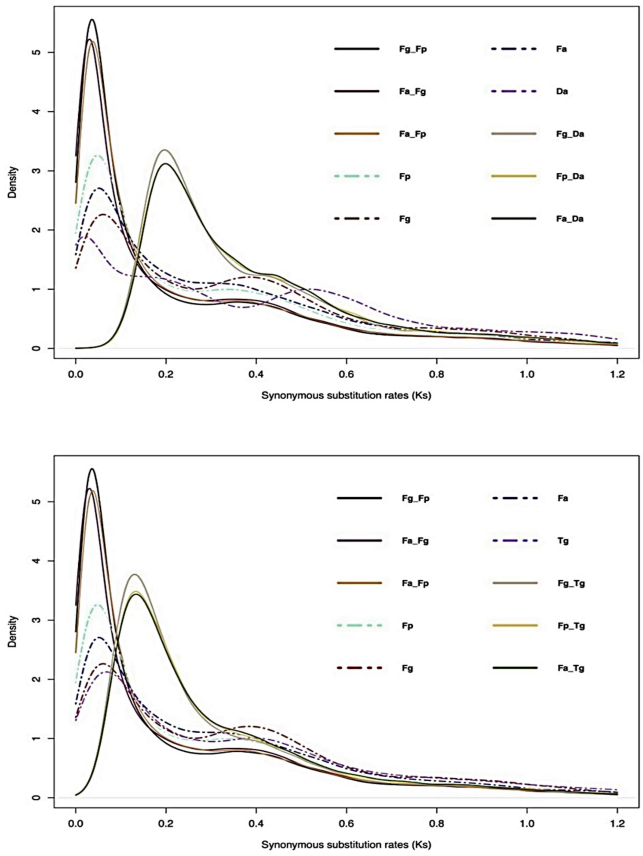

Figure 5. *Ks* distribution plots for paralog pairs in three *Ferula* species and orthologs among *Ferula*, *T. garganica,* and *D. carota*.(A). *Ferula* species were compared to *D. carota* (B), *Ferula* species were compared to *T. gargan*

detect whether WGD in *F. gummosa* had occurred ealier than the WGD in other species or not. In further investigation, the *Ferula* paralogs distribution indicated, WGD in this species was much older than the divergence of *Ferula* from *D. carota*. Our simulations confirmed findings from previous research that there were two specific WGDs in carrot (Iorizzo *et al*., 2016), which likely had occurred at 43 and 70 MYA, respectively.

Later, a phylogenetic tree was constructed in terms of 1,000 random single copy gene from all the five species (Figure 6). According to our phylogenetic results, all the three *Ferula* species were grouped together in one monophyletic clade, with the estimated divergence time of 2.99 MYA between *F. gummosa* and *F. assa-foetida,* nearly in agreement with the value computed in the Timetree website, and 3.87 MYA between *F. persica* and two other *Ferula* species. The results suggest that both *F. gummosa*, and *F. assa-foetida* are evolutionary closer to each other than *F. persica*, indicating that they have been possibly divergent from *F. persica*. Hence, they could be called as "newly divergent *Ferula* species" as compared to the *F. persica*, or suggesting that *F. persica* possibly possesses an "evolutionary older history" than the other two *Ferula* species. This observation was in close agreement with the earlier works (Kurzyna-Młynik *et al*., 2008), whose evidence clarified the fact that Mediterranean *Ferula* lineages originated from Asian ancestors, as well as the general theory of the westward colonization by Asian steppe plants (Franzke *et al*., 2004).

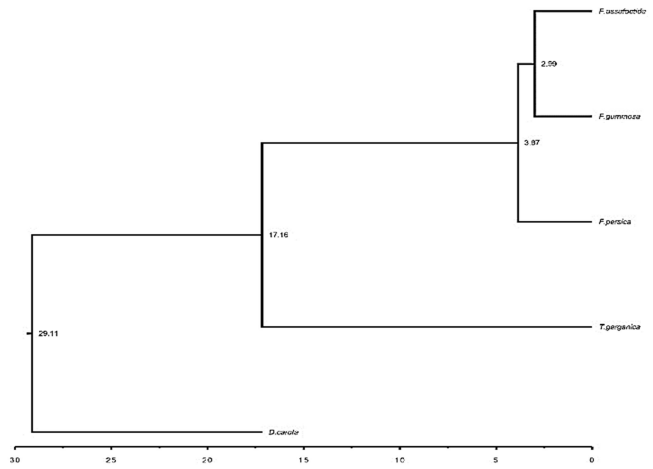

Figure 6. Phylogenetic relationships among *Ferula* species, *Thapsia* and *Daucus* inferred from 1000 single-copy orthogroups.

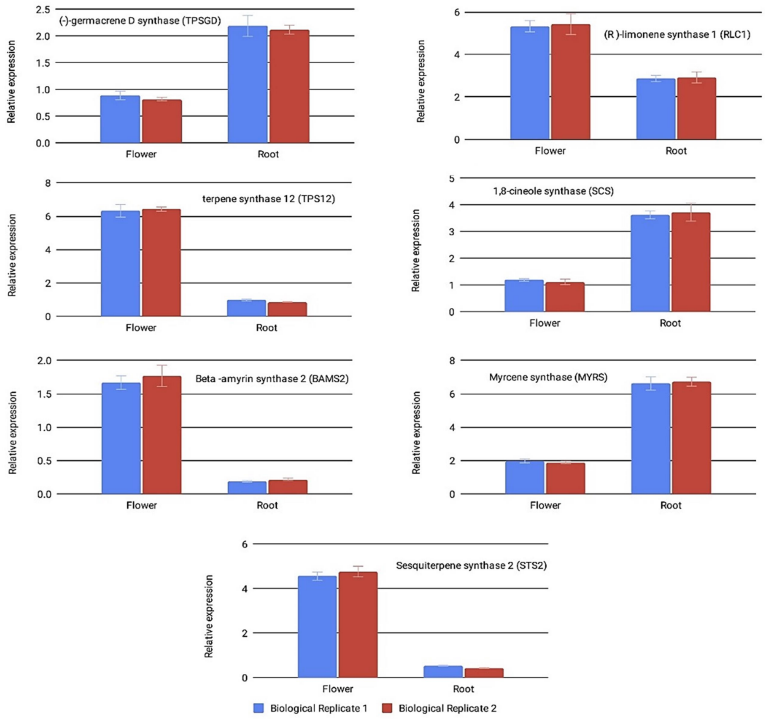

Figure 7. Gene expression patterns of the sleeted seven genes to verify RNA-Seq data.

Verification of RNA-Seq data by qRT-PCR

The expression level of seven candidate genes involved in the terpenoid biosynthesis pathway, including *TPS12*, *STS2*, *MYRS*, *BAMS2*, *SCS*, *RLC1*, and *TPSGD* were evaluated via three technical replicates for each one of the two biological replicates per tissues (Figure 7). Overall, qRT-PCR results confirmed the expression profiles detected by DEGs analysis from transcriptome data.

CONCLUSIONS

In conclusion, our results indicated that the production of secondary metabolites and expression patterns of the corresponding genes in *F. persica* could be a tissue-specific phenomenon. Considering GC-MS profiling, the monoterpene and sesquiterpene compounds exhibited the maximum and minimum levels in both tissues of flower and root. Based on DEGs, 396 and 1731 transcripts were up-regulated in, respectively, root and flower tissues, of which 86 were totally annotated as TPSs. Lastly, according to our phylogenetic results, all the three *Ferula* species formed only one monophyletic clade, with the estimated divergence time of 2.99 MYA between *F. gummosa* and *F. assa-foetida*, and 3.87 MYA between *F. persica* and two other *Ferula* species, indicating an older history for *F. persica* than that that of both *F. gummosa* and *F. assa-foetida*.

ACKNOWLEDGEMENTS

The Iran National Science Foundation (INSF, No. 4001289) as well as a visiting grant of 040.11.685 awarded by The Dutch Research Council (NWO) to MRN and AD financially supported this work.

REFERENCES

1. Amini, H. Naghavi, M. R., Shen, T. Wang, Y., Nasiri, J., Khan, I. A., Fiehn, O., Zerbe, P. and Maloof, J. N. 2019. Tissue-Specific Transcriptome Analysis Reveals Candidate Genes for Terpenoid and Phenylpropanoid Metabolism in the Medicinal Plant Ferula assafoetida. *G3: Genes Genom. Genet.,* **9:** 807-816.
2. Ashour, M., Wink, M. and Gershenzon, J. 2018. *Biochemistry of Terpenoids: Monoterpenes, Sesquiterpenes and Diterpenes.* Volume 40: Biochemistry of Plant Secondary Metabolism, Annual Plant Reviews Book Series, PP. 258-303.
3. Asili, J., Sahebkar, A., Fazly, B. S., Bazzaz, S., Sharifi, and Iranshahi, M. 2009. Identification of Essential Oil Components of Ferula badrakema Fruits by GC-MS and 13 C-NMR Methods and Evaluation of its Antimicrobial Activity. *J. Essent. Oil-Bear. Plants,* **12:** 7-15.
4. Barzegar, A., Salim, M. A., Badr, P., Khosravi, A., Hemmati, S., Seradj, H., Iranshahi, M. and Mohagheghzadeh, A. 2020. Persian Asafoetida vs. Sagapenum: Challenges and Opportunities. *Res. J. Pharmacogn.,* **7:** 71-80.
5. Bolger, A. M.,Lohse, M. and Usadel, B. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics,* **30:** 2114-2120.
6. Capella-Gutiérrez, S., Silla-Martínez, J. M. and Gabaldón, T. 2009. TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses. *Bioinformatics,* **25:** 1972-1973.
7. Emms, D. M. and Kelly, S. 2015. OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy. *Genome Biol.,* **16:** 157.
8. Franzke, A., Hurka, H., Janssen, D., Neuffer, B., Friesen, N., Markov, M. and Mummenhoff, K. 2004. Molecular Signals for Late Tertiary/Early Quaternary Range Splits of an Eurasian Steppe Plant: *Clausia aprica* (Brassicaceae). *Mol. Ecol,* **13:** 2789-2795.
9. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E.,

Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. 2011. Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome. *Nat. Biotechnol.,* 29: 644.

10. Hong, G. -J., Xue, X. -Y., Mao, Y. -B., Wang, L. -J. and Chen, X. -Y. 2012. Arabidopsis MYC2 Interacts with DELLA Proteins in Regulating Sesquiterpene Synthase Gene Expression. *Plant Cell,* **24:** 2635-2648.
11. Iranshahi, M., Amin, G., Sourmaghi, M. S., Shafiee, A. and Hadjiakhoondi, A. 2006. Sulphur-Containing Compounds in the Essential Oil of the Root of *Ferula persica* Willd. var. Persica. *Flavour. Fragr. J.,* **21**: 260-261.
12. Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., Bowman, M., Iovene, M., Sanseverino, W., Cavagnaro, P., Yildiz, M., Macko-Podgórni, A., Moranska, E., Grzebelus, E., Grzebelus, D., Ashrafi, H., Zheng, Z., Cheng, S., Spooner, D., Van Deynze, A. and Simon, P. 2016. A High-Quality Carrot Genome Assembly Provides New Insights into Carotenoid Accumulation and Asterid Genome Evolution. *Nat. Genet.,* **48:** 657-666.
13. Javidnia, K., Miri, R., Kamalinejad, M. and Edraki, N. 2005. Chemical Composition of *Ferula persica* Wild. Essential Oil from Iran. *Flavour. Fragr. J.,* **20:** 605-606.
14. Ji, Y., Xiao, J., Shen, Y., Ma, D., Li, Z., Pu, G., Li, X., Huang, L., Liu, B. and Ye, H., Wang, H. 2014. Cloning and Characterization of AabHLH1, a bHLH Transcription Factor that Positively Regulates Artemisinin Biosynthesis in *Artemisia annua*. *Plant Cell Physiol.,* **55:** 1592-1604.
15. Jimenez-Jacinto, V., Sanchez-Flores, A. and Vega-Alvarado, L. 2019. Integrative Differential Expression Analysis for Multiple EXperiments (IDEAMEX): A Web Server Tool for Integrated RNA-seq Data Analysis. *Front. Genet.,* **10**: 279.
16. Kavoosi, G. and Rowshan, V. 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.
17. Kurzyna-Młynik, R., Oskolski, A. A., Downie, S. R., Kopacz, R., Wojewódzka, A. and Spalik, K. 2008. Phylogenetic Position of the Genus *Ferula* (Apiaceae) and Its Placement in Tribe Scandiceae as Inferred from nrDNA ITS Sequence Variation. *Plant Syst. Evol.,* **274**: 47.
18. Li, B. and Dewey, C. N. 2011. RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without aReference Genome. *BMC Bioinform.,* **12:** 323.
19. Livak, K. J. and Schmittgen, T. D. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods,* 25: 402-408.
20. Love, M., Huber, I. W. and Anders, S. 2014. Moderated Estimation of Fold Change and Dispersion for RNA-seq Data with DESeq2. *Genome Biol.,* **15**: 550.
21. Mabberley, D. J. 2017. *Mabberley's Plant-Book: A Portable Dictionary of Plants, Their Classification and Uses*. 4 Edition. Cambridge: Cambridge University Press.
22. Najafabadi, A.S., Naghavi, M. R., Farahmand, H. and Abbasi, A. 2017.

Transcriptome and Metabolome Analysis of *Ferula gummosa* Boiss. to Reveal Major Biosynthetic Pathways of Galbanum Compounds. *Funct. Integr. Genomics,* **17:** 725-737.

23. Nasiri, J., Motamedi, E., Naghavi, M. R. and Ghafoori, M. 2019. Removal of Crystal Violet from Water Using β-Cyclodextrin Functionalized Biogenic Zero-Valent Iron Nanoadsorbents Synthesized via Aqueous Root Extracts of *Ferula persica*. *J. Hazard. Mater.*, **367:** 325-338.
24. Nasiri, J. Rahimi, M. Hamezadeh, Z. Motamedi, E. and Naghavi, M. R. 2018. Fulfillment of Green Chemistry for Synthesis of Silver Nanoparticles Using Root and Leaf Extracts of *Ferula persica*: Solid-State Route vs. Solution-Phase Method. *J. Clean. Prod.,* **192:** 514-530.
25. Priya, P., Yadav, A., Chand, J. and Yadav, G. 2018. Terzyme: A Tool for Identification and Analysis of the Plant Terpenome. *Plant Methods,* **14:** 4-4.
26. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. 2015. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.,* 43: e47-e47.
27. Robinson, M. D., McCarthy, D. J. and Smyth, G. K. 2010. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics,* **26:** 139-140.
28. 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.
29. Singh B. and Sharma, R. A. 2015. Plant Terpenes: Defense Responses, Phylogenetic Analysis, Regulation and Clinical Applications. *3 Biotech,* **5:** 129-151.
30. Soorni, A., Borna, T., Alemardan, A., Chakrabarti, M., Hunt, A. G. and Bombarely, A. 2019. Transcriptome Landscape Variation in the Genus *Thymus*. *Genes,* **10:** 620.
31. Spyropoulou, E. A., Haring, M. A. and Schuurink, R. C. 2014. RNA Sequencing on *Solanum lycopersicum* Trichomes Identifies Transcription Factors that Activate Terpene Synthase Promoters. *BMC Genom.,* **15**: 402.
32. Tarazona, S., García, F., Ferrer, A., Dopazo, J. and Conesa, A. 2011. NOIseq: A RNA-seq Differential Expression Method Robust for Sequencing Depth Biases, *EMBnet. J.,* **17:** 18-19.
33. Tian, T. Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W. and Su, Z. 2017. AgriGO v2.0: A GO Analysis Toolkit for the Agricultural Community, 2017 Update. *Nucleic Acids Res.,* **45:** W122-W129.
34. Wang S. and Gribskov, M. 2017. Comprehensive Evaluation of de Novo Transcriptome Assembly Programs and Their Effects on Differential Gene Expression Analysis. *Bioinformatics,* **33:** 327-333.
35. Xu, Y. -H., Wang, J. -W., Wang, S., Wang, J. -Y. and Chen, X. -Y. 2004. Characterization of GaWRKY1, a Cotton Transcription Factor that Regulates the Sesquiterpene Synthase Gene (+)-δ-Cadinene Synthase-A. *Plant Physiol.,* **135**: 507-515.

36. Yamada Y. and Sato, F. 2013. Transcription Factors in Alkaloid Biosynthesis. In: "*International Review of Cell and Molecular Biology*". Ed: Elsevier, **305:** 339-382.
37. Yokotani, N., Sato, Y., Tanabe, S., Chujo, T., Shimizu, T., Okada, K., Yamane, H., Shimono, M., Sugano, S., Takatsuji, H., Kaku, H., Minami, E. and Nishizawa, Y. 2013. WRKY76 Is a Rice Transcriptional Repressor Playing Opposite Roles in Blast Disease Resistance and Cold Stress Tolerance. *J. Exp. Bot.,* **64(16)**: 5085-5097.
38. Yoshida, K. , Oyama-Okubo, N. and Yamagishi, M., 2018. An R2R3-MYB Transcription Factor ODORANT1 Regulates Fragrance Biosynthesis in Lilies (*Lilium* spp.). *Mol. Breed.,* **38:** 144.
39. Zhang, M., Jin, X., Chen, Y., Wei, M., Liao, W., Zhao, S. Fu, C. and Yu, L. 2018. TcMYC2a, a Basic Helix–Loop–Helix Transcription Factor, Transduces JA-Signals and Regulates Taxol Biosynthesis in *Taxus chinensis. Front. Plant Sci.*, **9:** 1-13.
40. Zvi, M. M. B., Shklarman, E., Masci, T., Kalev, H., Debener, T., Shafir, S. Ovadis, M. and Vainstein, A. 2012. PAP1 Transcription Factor Enhances Production of Phenylpropanoid and Terpenoid Scent Compounds in Rose Flowers. *New Phytol.,* **195(2):** 335-345

مسیر بیوسنتزی ترپنوئید در *Ferula persica* با استفاده از آنالیز رونوشت و داده های متابولوم

ج. نصیری، ا. سورنی، آ. د. ج. وان دایک، و م. ر. نقوی

چکیده

در این تحقیق، آنالیز پروفایل متابولوم و ترانسکریپتوم گیاه کما (*Ferula persica*) از طریق داده‌های GC-MS و RNA-seq انجام شد. آنالیز اسانس‌های استخراج شده از هر دو بافت گل و ریشه، برتری کمّی ترکیبات مونوترپن را نشان داد، درحالی‌که ترکیبات سسکوئی‌ترپن در مقادیر کمتری وجود داشتند. برمبنای آنالیز RNA-seq، حدود ۲۱۲۷ ژن با بیان متفاوت بین ریشه (۳۹۶ رونوشت با بیان بالا) و گل (۱۷۳۱ رونوشت با بیان بالا) یافت شد. از ۲۱۲۷ رونوشت، ۸۶ رونوشت بعنوان ترپن‌سینتاز (TPSs) دسته‌بندی شدند، که ۸۳ رونوشت معادل TPS متعاقباً در پنج زیرخانواده جداگانه TPS-a (33)، TPS-b (42)، TPS-c (2)، TPSe-f (3) و TPS-g (3) طبقه‌بندی شدند. چندین خانواده از فاکتورهای رونویسی در بین ژن‌های با بیان متفاوت شناسایی شدند که نقش تنظیمی مستقیم یا غیرمستقیم آنها را برای بیوسنتز ترپنوئیدها در *F. persica* نشان می‌دهد. درنهایت، با توجه به نتایج فیلوژنی، هر دوگونه *F. assa-foetida* و *F. gummosa* در یک کلاستر قرار گرفتند، درحالی‌که *F. persica* در یک کلاستر مجزا مستقر شد، ضمن اینکه زمان واگرایی

دوگونه *F. assa-foetida* و*F. gummosa* حدود ۲/۹۹ میلیون سال پیش، و بین *F. persica* و دو گونه دیگر جنس فرولا حدود ۳/۹۸ میلیون سال پیش تخمین زده شد.