



Article

Full-Length Transcriptomic Sequencing and Temporal Transcriptome Expression Profiling Analyses Offer Insights into Terpenoid Biosynthesis in *Artemisia argyi*

Ran Xu ^{1,*}, Yue Ming ¹, Yongchang Li ² , Shaoting Li ¹, Wenjun Zhu ¹, Hongxun Wang ¹, Jie Guo ³, Zhaohua Shi ⁴, Shaohua Shu ⁵ , Chao Xiong ¹, Xiang Cheng ¹, Limei Wang ¹, Jingmao You ^{3,*} and Dingrong Wan ^{6,*}

¹ School of Life Science and Technology, Wuhan Polytechnic University, Wuhan 430023, China

² College of Osteopathic Medicine, Kansas City University of Medicine and Biosciences, Joplin, MO 64804, USA

³ Institute of Chinese Herbal Medicines, Hubei Academy of Agricultural Sciences, Enshi 445000, China

⁴ Key Laboratory of Chinese Medicine Resources and Compound Formula, Ministry of Education, Hubei University of Chinese Medicine, Wuhan 430065, China

⁵ College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

⁶ School of Pharmaceutical Sciences, South-Central University for Nationalities, Wuhan 430074, China

* Correspondence: rxu@whpu.edu.cn (R.X.); jingmaoyou@126.com (J.Y.); wandr666@163.com (D.W.)

Abstract: *Artemisiae argyi* Folium is a traditional herbal medicine used for moxibustion heat therapy in China. The volatile oils in *A. argyi* leaves are closely related to its medicinal value. Records suggest that the levels of these terpenoids components within the leaves vary as a function of harvest time, with June being the optimal time for *A. argyi* harvesting, owing to the high levels of active ingredients during this month. However, the molecular mechanisms governing terpenoid biosynthesis and the time-dependent changes in this activity remain unclear. In this study, GC–MS analysis revealed that volatile oil levels varied across four different harvest months (April, May, June, and July) in *A. argyi* leaves, and the primarily terpenoids components (including both monoterpenes and sesquiterpenes) reached peak levels in early June. Through single-molecule real-time (SMRT) sequencing, corrected by Illumina RNA-sequencing (RNA-Seq), 44 full-length transcripts potentially involved in terpenoid biosynthesis were identified in this study. Differentially expressed genes (DEGs) exhibiting time-dependent expression patterns were divided into 12 coexpression clusters. Integrated chemical and transcriptomic analyses revealed distinct time-specific transcriptomic patterns associated with terpenoid biosynthesis. Subsequent hierarchical clustering and correlation analyses ultimately identified six transcripts that were closely linked to the production of these two types of terpenoid within *A. argyi* leaves, revealing that the structural diversity of terpenoid is related to the generation of the diverse terpene skeletons by prenyltransferase (TPS) family of enzymes. These findings can guide further studies of the molecular mechanisms underlying the quality of *A. argyi* leaves, aiding in the selection of optimal timing for harvests of *A. argyi*.

Keywords: *Artemisia argyi*; SMRT sequencing; temporal transcriptome; gene expression patterns; terpenoid biosynthesis



Citation: Xu, R.; Ming, Y.; Li, Y.; Li, S.; Zhu, W.; Wang, H.; Guo, J.; Shi, Z.; Shu, S.; Xiong, C.; et al. Full-Length Transcriptomic Sequencing and Temporal Transcriptome Expression Profiling Analyses Offer Insights into Terpenoid Biosynthesis in *Artemisia argyi*. *Molecules* **2022**, *27*, 5948. <https://doi.org/10.3390/molecules27185948>

Academic Editors: Changsheng Zhang and Nurhayat Tabanca

Received: 22 May 2022

Accepted: 31 August 2022

Published: 13 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artemisia argyi is among the most common *Artemisia* species, growing as a native plant species in China, Japan and the Korea. The leaves of *A. argyi* are often used in the traditional Chinese medicine (TCM) practice of moxibustion, for the treatment of diarrhea, tuberculosis, eczema, hemostasis, and menstruation-associated symptoms [1]. The dried leaves of *A. argyi* are often used as a food ingredient, owing to their delicious flavor and characteristic smell [2–4].

“Chinese Mugwort” (*A. argyi*) is the most widely used medicinal species in China. Owing to its importance as a traditional medicinal plant, ancient Chinese customs suggest

that its buds and leaves be harvested before and after 5 May of the lunar calendar (usually in early June), respectively, for medicinal use. Recent analyses have similarly shown that June may be the optimal time for *A. argyi* harvesting, owing to the high levels of pharmacologically active volatile oils, tannins, and phenolics within plant tissues at this time point, which may be associated with improved antitumor, antiviral, antimicrobial, anti-inflammatory, and immunomodulatory properties [5]. In addition, dried and ground *A. argyi* leaves are the original material for moxa floss, which is used for moxibustion as a TCM therapeutic to cure dysmenorrhea, diarrhea, and fatigue, and the *A. argyi* leaf's volatile oils play a significant therapeutic role in moxibustion [6,7]. So, the total volatile oil content in *A. argyi* samples is often assessed as a measure of the quality of this herb, given that these oils are the major active ingredients.

The volatile oils derived from *A. argyi* leaves are primarily composed of monoterpenes and sesquiterpenes [8,9], both of which are members of a large and structurally diverse terpenoid family derived from conjugations of five-carbon dimethylallyl diphosphate (DMAPP) and its isomer isopentenyl diphosphate (IPP). Prenyl transferases are enzymes responsible for catalyzing DMAPP condensation with multiple units of IPP in a head–tail orientation, yielding prenyl diphosphate chains of varying lengths. Terpene synthases (TPSs) and cytochrome P450 (CYP) family members subsequently rearrange these compounds and drive terpenoid diversification, leading to the production of thousands of different isoprenoids. For example, monoterpene synthase and sesquiterpene synthase can generate monoterpenes and sesquiterpenes from geranyl diphosphate (GPP) and farnesyl diphosphate (FPP) precursors [10,11], respectively.

While the levels of these different terpenoids within *A. argyi* leaves vary as a function of harvest time, the molecular basis for these components' time-dependent changes remains elusive. As such, the present study employed a gas chromatography–mass spectrometry (GC–MS) approach to assess differences in the *A. argyi* leaf's volatile-oil-accumulation patterns during different harvest months (April, May, June, and July). This analysis was conducted in parallel with single-molecule real-time (SMRT) sequencing and Illumina RNA-sequencing (RNA-Seq) analyses of *A. argyi* leaf samples collected at these different harvest time points, in order to clarify the underlying transcriptomic profiles in these leaves. The underlying metabolic pathways associated with the terpene biosynthesis in these samples were identified through Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses, and the differentially expressed genes (DEGs) associated with monoterpene and sesquiterpene biosynthesis during different harvest months were further evaluated. We additionally conducted preliminary screening for key terpenoid-biosynthesis-related candidate genes, through Pearson's correlation analyses of the total volatile oil contents and gene expression patterns in different leaf samples. Therefore, this study provides a robust foundation for future studies of terpenoid synthesis in *A. argyi* leaves, potentially guiding genetic alteration efforts aimed at improving the yield of this culturally and economically important medicinal plant species.

2. Results

2.1. Assessment of *A. Argyi* Leaf Volatile Oil Contents As a Function of Harvest Time

An initial GC–MS analysis of volatile compounds identified 60 compounds from an *A. argyi* leaf sample (Table S1), corresponding to 70% of the volatile oil content and consisting of a complex mixture of monoterpenes (58.3%), sesquiterpenes (28.3%), and other components. Ten volatile oil compounds were identified in *A. argyi*. The primary monoterpenes included *O*-Cymene; 1,8-Cineole; (1*R*)-4-methyl-1-propan; γ -Terpinene; Sabinene; Thujone; and α -Terpineol, and the primary sesquiterpenes compounds were β -Caryophyllene, β -Farnesene, and β -Selinene (Figure 1A,B and Figure S1).

We further compared the levels of these volatile oil components in *A. argyi* leaves as a function of harvest time, by selecting three replicate samples at four time points from April to July in the same year. The results indicated that monoterpene and sesquiterpene levels varied over time in the analyzed *A. argyi* leaf samples. Six of the highest volatile oil

levels in *A. argyi* leaves were from samples harvested in June, while two were observed in April and one each was observed in May and July. Total levels of monoterpenes and sesquiterpenes initially rose to peak levels of 23.1% and 8.2% in early June, respectively, before falling by about 17.7% and 6.2% in July (Figure 1C,D). This suggested that the volatile oil accumulation was closely associated with the developmental stage of individual *A. argyi* plants.

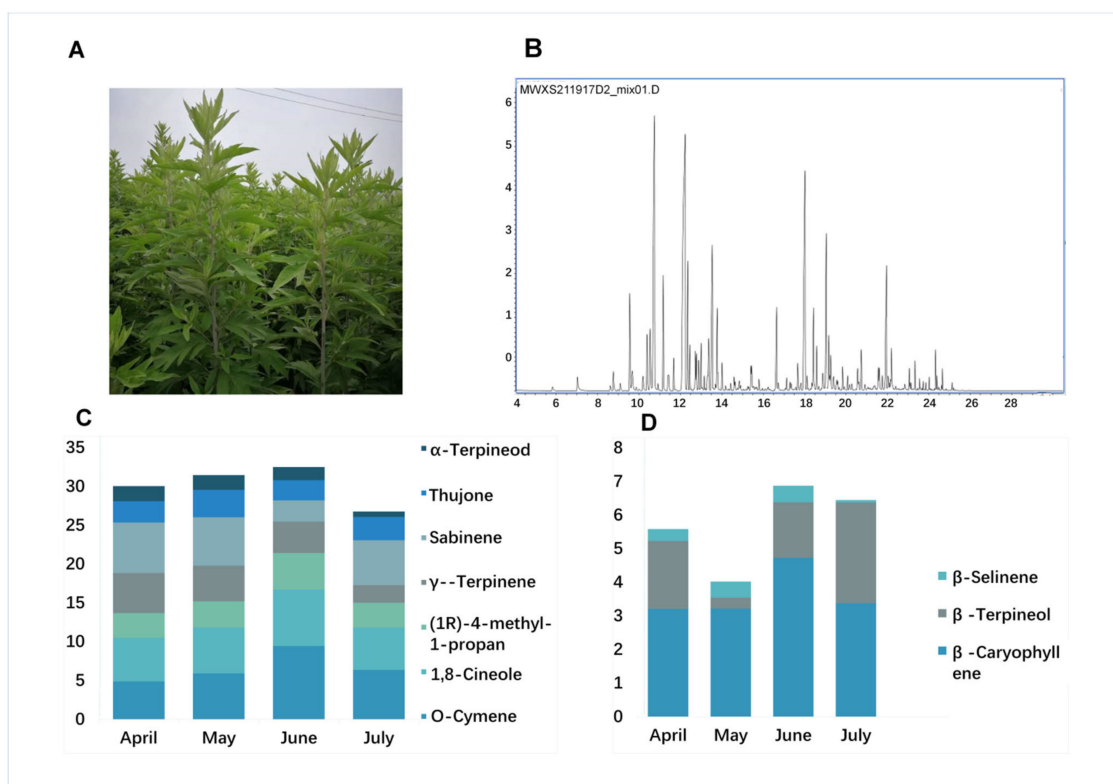


Figure 1. GC–MS chromatograms and terpenoid contents in *A. argyi*. (A) *A. argyi* plants cultivated under field conditions. (B) GC–MS total ion chromatograms of volatile compounds from *A. argyi* leaf samples collected in April, May, June, and July, respectively. (C) Bar graph corresponding to the distribution of levels of seven monoterpenes in *A. argyi* samples collected at four time points. (D) Bar graph corresponding to the distribution of levels of three sesquiterpenes in *A. argyi* samples collected at four time points.

2.2. SMRT and Illumina RNA-Seq Analyses of *A. Argyi* Leaf Samples

Next, we performed RNA-Seq-based transcriptomic profiling to generate a de novo transcriptomic assembly. A representative full-length transcriptome was generated by pooling equimolar amounts of DNA from four leaf samples from different collection months, prior to SMRT sequencing. A Pacific RSII sequencing instrument was used to sequence differently sized full-length cDNA libraries (1–2 kb, 2–3 kb, and 3–6 kb), ultimately generating 63,281,536 subreads (84.81 Gb) in one SMRT cell with an average read length of 1267 bp and an N50 of 1571 bp (Tables S2 and S3). In total, 888,748 read of inserts (ROIs) were generated, among which 855,813 harbored two primers and a poly-A tail and were, thus, considered to be full-length ROIs. These were further subdivided, with 740,702 being identified as full-length non-chimeric (FLNC) reads at an average read length of 1268 bp (Figure S2A,B and Tables S4 and S5).

As SMRT sequencing can have high error rates, iterative clustering for error correction (ICE) analyses were conducted, with transcripts being clustered together to predict consensus full-length non-chimeric read of insert (FLNC-ROI) isoforms. In total, 241,963 consensus isoform sequences were identified with different sizes (0–1 kb (24,306), 1–2 kb (33,989), 2–3 kb (9304), and > 3 kb (1579)), and the mean length of these sequences

was 1395 bp (Table S6). Next, cDNA libraries were prepared from the same samples utilized for SMRT sequencing, with a deep RNA-seq being performed with an Illumina HiSeq X-ten instrument. This analysis yielded 103.04 Gb of clean reads that were used to further correct SMRT transcripts. Following error correction, 69,178 unigenes with an N50 length of 1619 bp were successfully identified from the 241,963 mRNAs, of which 46,946 unigenes were longer than 2000 bp. Relative to the assembled unigenes obtained in prior NGS analyses of *A. argyi* [11], the number of long-length unigenes in our dataset was greatly increased. Then, using BLAST, the SwissProt database was compared with Hmmscan to search for Pfam domain homology as a means of predicting coding regions. A total of 69,178 complete transcripts were obtained (Tables S7 and S8). The details of the prediction results are listed in Table S9, and the coding sequence (CDS) length distributions are shown in Figure S2C.

2.3. Functional Annotation

To more fully understand the mechanisms governing terpenoid biosynthesis in *A. argyi*, functional annotation of FLNC transcripts was conducted using a series of databases. In total, 69,178 unique isoforms were annotated with BLASTX (v 2.2.26) and the COG (Clusters of Orthologous Groups) [12], GO [13], KEGG [14], KOG [15], Pfam (Protein family) [16], Swissprot [17], and NR [18] databases (Figure 2A, Table S10).

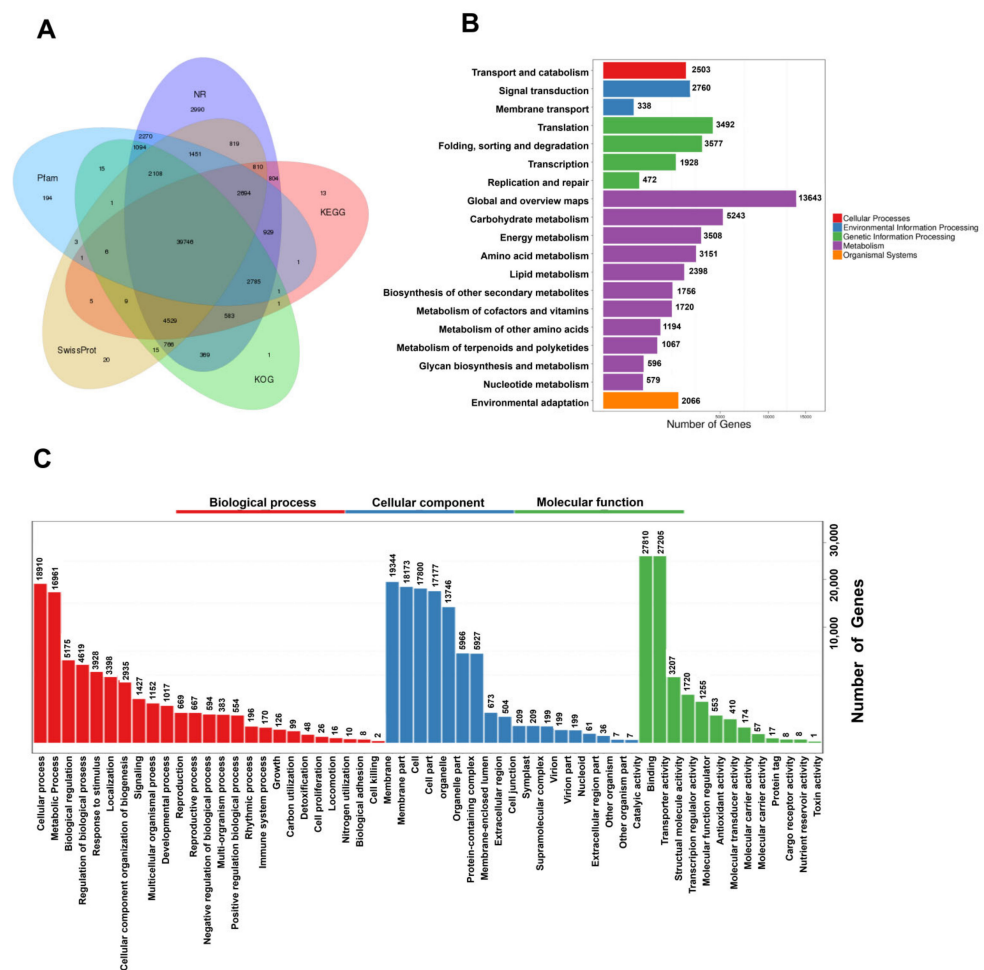


Figure 2. Functional annotation and classification of assembled *A. argyi* unigenes. (A) Annotation details arranged in a Venn diagram. (B) KEGG pathway classifications for putative proteins. (C) GO classifications for *A. argyi* unigenes, including biological process (BP), cellular component (CC), and molecular function (MF) annotations.

GO functional annotations were used to assign specific biological process (BP), molecular function (MF), and cellular component (CC) terms to the unique isoforms, yielding three primary GO categories and 56 subcategories. A large proportion of these genes were annotated in the BP terms (“cellular process”, “metabolic process”, and “biological regulation”), the CC terms (“membrane”, “membrane part”, and “cell”), and the MF terms (“catalytic activity”, “binding”, and “transcription regulator activity”) (Figure 3C).

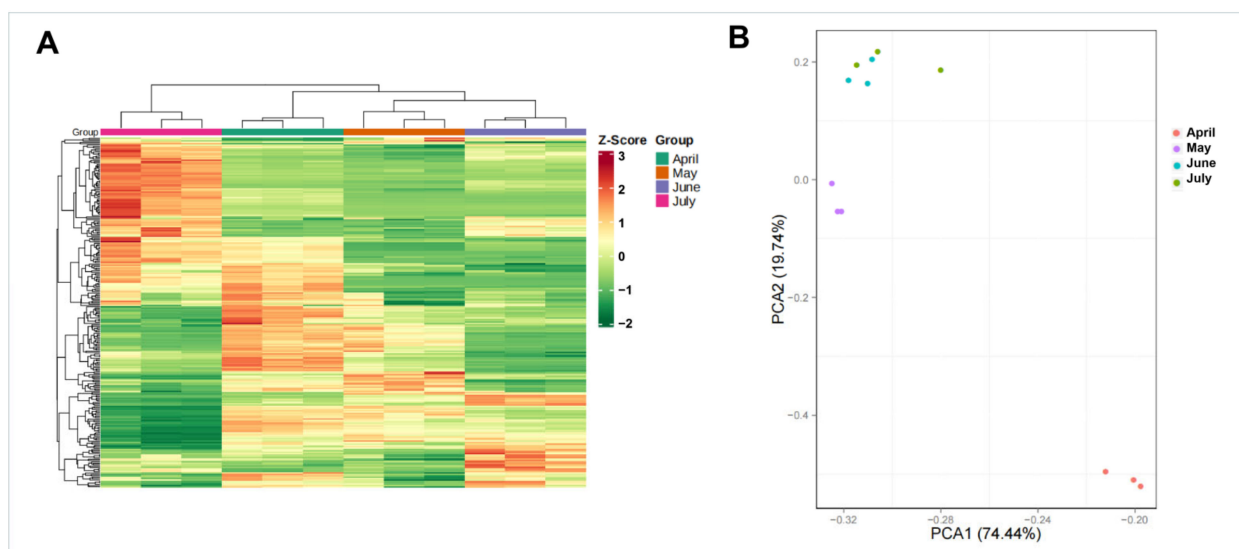


Figure 3. Gene expression levels during four *A. argyi* developmental stages. (A) Clustering dendrogram of gene expression patterns in different *A. argyi* developmental stage samples. (B) PCA of gene expression profiles for different *A. argyi* developmental stage samples.

In total, these genes were associated with 137 KEGG pathways, many of which were related to growth and metabolite accumulation (Figure 2B).

2.4. Individual Transcriptomic Analyses of *A. Argyi* Leaf Samples Collected at Four Different Developmental Time Points

Next, cluster dendrogram and principal component analysis (PCA) analyses were conducted using samples from different *A. argyi* developmental stages. Samples from these four *A. argyi* developmental stages were clustered into four separate groups, and the samples from the same collection month were assembled into the same branches, based on the results of the clustering tree, revealing clear differences in gene expression profiles among these samples (Figure 3A). PCA results revealed that the samples from May and April were significantly different from those collected in June and July, based upon the PC1 axis (Figure 3B), which is consistent with the results observed in the clustering dendrogram.

The time-specific expression of genes was further assessed to reveal differences in the transcripts from the *A. argyi* samples from four different time points. In total, 67,117 unigenes were applied to create gene intersection profiles. A total of 52,867 unigenes were shared among *A. argyi* from the four harvest time points, with 1006, 658, 536, and 616 unigenes being specifically expressed in April, May, June, and July, respectively (Figure 4).

2.5. Identification and Enrichment Analyses of DEGs Associated with Particular *A. Argyi* Developmental Stages

To further explore patterns of differential gene expression across these *A. argyi* samples collected at the four selected harvest time points, we employed fold change (FC) and FDR values to compare the expression among samples [11], leading to the identification of 140,611 DEGs, with individual comparisons among samples being made to identify DEGs that were up- or down-regulated. Specifically, DEGs were identified by comparing samples

from May and April, April and June, April and July, May and June, May and July, and June and July. The greatest number of DEGs (33,968) was observed when comparing samples from April and July, while the smallest number (15,351) was evident when comparing samples from June and July (Figure S3), consistent with the important roles of many of these genes during the different stages of *A. argyi* development.

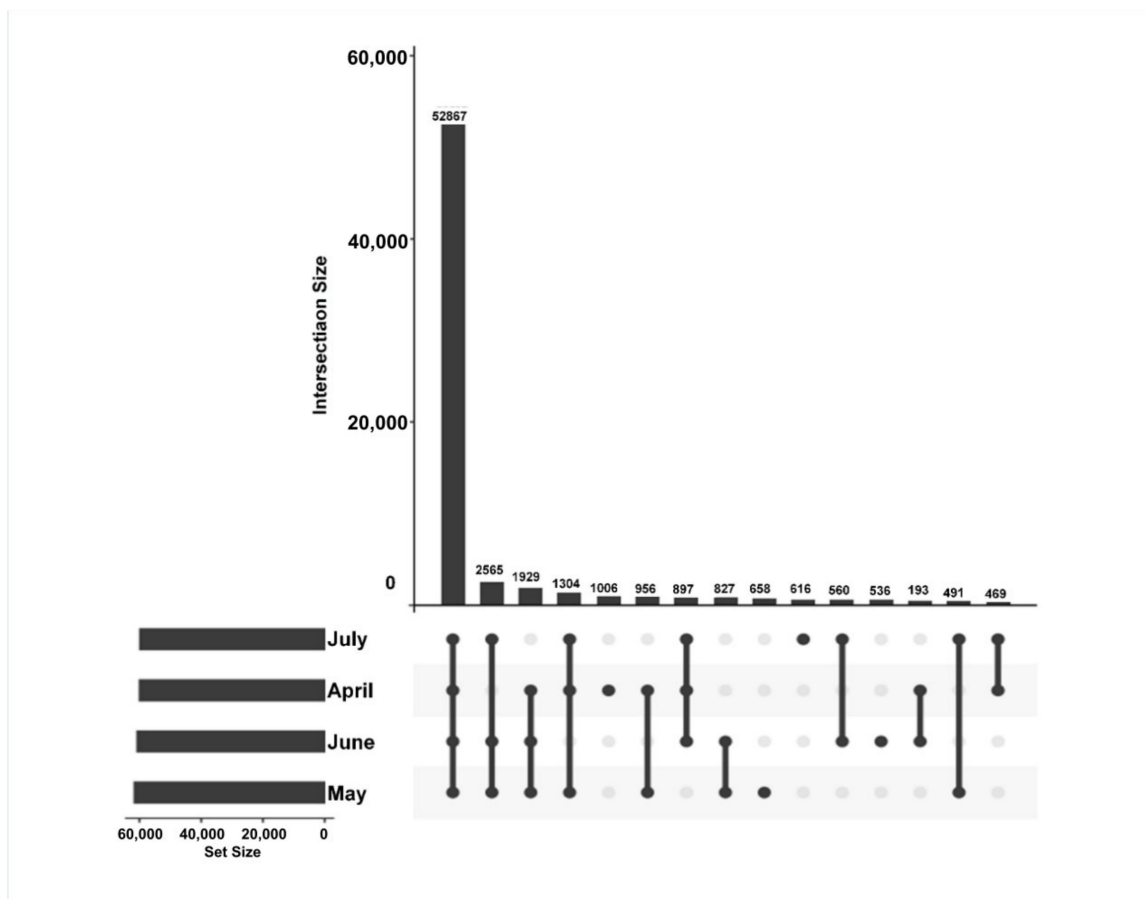


Figure 4. Gene intersection matrix profiles for *A. argyi* samples collected at four time points.

In a soft clustering analysis, 12 gene clusters were obtained with distinctive characteristics. For example, clusters 10 and 12 exhibited trends toward consistent up-regulation and down-regulation over time, respectively. The other 10 clusters exhibited varying patterns of expression changes over time from April through July. Notably, gene expression in clusters 3 and 5 tended to rise from April to June and to decrease in July; these genes, thus, exhibited trends consistent with the overall trend in volatile oil accumulation at these harvest time points (Figure 5). As such, these two clustering modules were considered to be correlated with *A. argyi* terpenoid synthesis.

2.6. Identification of Terpenoid-Biosynthesis-Related Genes

Through the integrated annotation approach outlined above, 77 unigenes were annotated as being associated with terpenoid backbone biosynthesis, suggesting that similarity-based BLASTX searches conducted with the Swiss-Prot database were more powerful than other tested analytical approaches. In order to assess the presence of these transcripts in other species, we conducted BLASTX analyses against a collection of plant cDNAs from protein-coding genes. We were able to identify 68 (88.3%) transcripts with at least one significant match, 17 of which exhibited high similarity to *Artemisia annua*. All unigenes with >500 bp in length, meeting functional annotation criteria, were ultimately evaluated,

leading to the identification of 77 terpenoid biosynthesis-related unigenes, 44 of which were full-length cDNA sequences (Table S11).

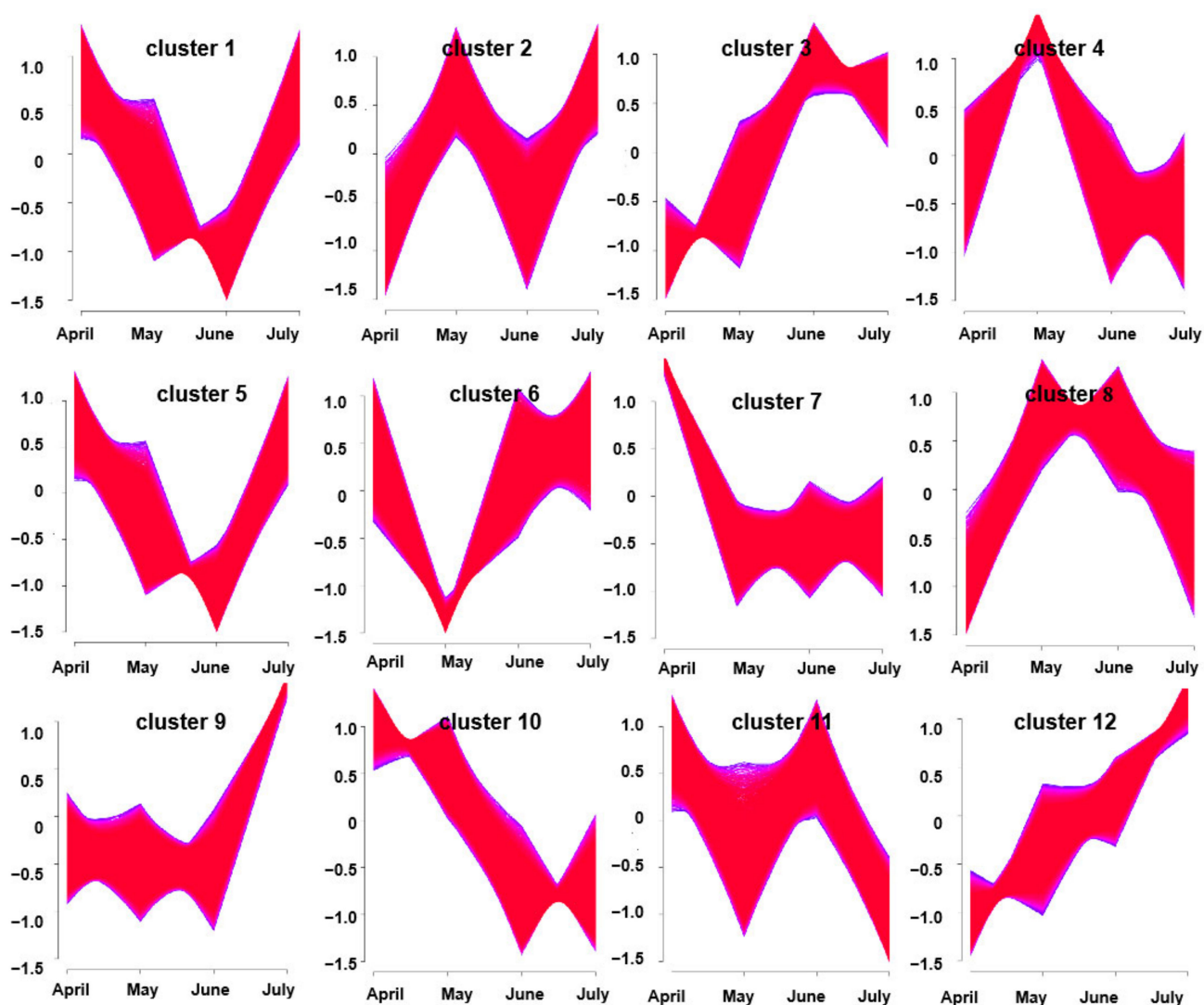


Figure 5. Clusters of DEGs obtained via soft clustering for *A. argyi* samples collected at four time points.

To clarify the potential biological functions of these genes, hierarchical clustering analyses were next performed based upon their log₂-transformed FPKM values in the analyzed *A. argyi* leaf samples. Ultimately, these genes were assigned to four distinct clusters. Group I contained 11 genes, including 1 gene in the MVA pathway (HMGR), 5 genes in the MEP pathway (1 HDS, 2 DXR, and 2 HDR), and 5 prenyltransferases (4 FPS, and 1 FPPS). These group I genes were expressed at the highest levels in *A. argyi* samples collected in May. In total, 29 genes were assigned to group II, including 9 genes in the MVA pathway (2 AACT, 3 MVK, 2 MDD, 1 PMK, and 1 HMDS), 15 genes in the MEP pathway (1 CMK, 9 DHR, 1 MVK, 2 PMK, 1 HDS, and 1 MCT), and 5 prenyltransferases (5 FPPS). These group II genes were expressed at the highest levels in samples collected in June or July. Group III contained 1 HMGS gene in the MVA pathway and 10 MEP pathway genes (2 HDS, 3 CMT, 1 CMK, 3 MDS, and 1 HDR) that were expressed at higher levels in the April or July samples. Lastly, 26 genes were assigned to group IV, including 1 HMGR gene in the MVA pathway, 15 MEP pathway genes (6 HDS, 2 MCT, 1 DXR, 5 MDS, and 1 HDR), and 10 prenyltransferases (2 IPPI, 2 FPPS, and 6 GPPS) that were expressed at the highest levels in the April samples (Figure 6 and Table S11).

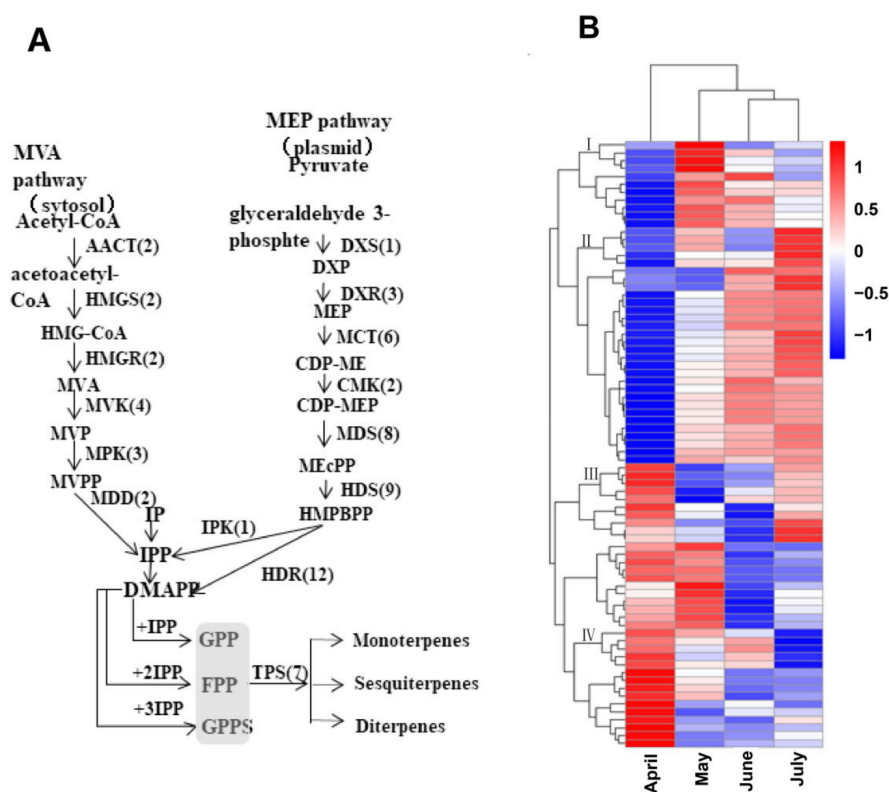


Figure 6. Schematic overview of the putative *A. argyi* terpenoid biosynthetic pathway, with corresponding gene expression levels. (A) The predicted pathway of *A. argyi* terpenoid biosynthesis, with the numbers in brackets denoting the number of enzymes in each family. (B) A clustered heatmap demonstrating the log₂-transformed FPKM expression levels of 77 terpenoid backbone genes and 7 TPS genes in *A. argyi* samples collected during the indicated months.

2.7. Assessment of Correlations between Terpenoid Content and Gene Expression

To better understand the regulation of volatile oil biosynthesis, transport, and accumulation over time in *A. argyi* leaves, we conducted further analyses of the genes associated with terpenoid backbone biosynthesis and modification. Eleven of these genes (from Figure 6) were maximally expressed in June, when these leaf samples exhibited peak volatile oil levels. Pearson correlation analyses revealed these genes to be significantly correlated with the terpenoid content in these leaf samples (Figure 7). In particular, the expression of HMGR1 from the MAV pathway was significantly correlated with 1,8-Cineole, (1R)-4-methyl-1-propan, γ -Terpinene, Sabinene, Thujone, and β -Selinene contents, while MDD1 and MDD2 were correlated with α -Terpineol levels. The HDR1 and HDR3 genes of the MEP pathway were also significantly correlated with O-Cymene, 1,8-Cineole, and β -Farnesene content, while HDS7 was significantly correlated with the γ -Terpinene, Sabinene, and β -Selinene levels. In addition, TPS2 was significantly positively correlated with the levels of (1R)-4-methyl-1-propan contents, and TPS5 was significantly positively correlated with the levels of the O-Cymene, Thujone, and β -Farnesene contents, respectively. These results identified six key genes: one 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR1), one (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (HDS7), two 1-deoxy-D-xylulose-5-phosphate reductoisomerase (HDR1, HDR7), and two terpene synthase (TPS2, TPS5) genes that were significantly correlated with the levels of monoterpenes or sesquiterpenes in the *A. argyi* samples. The structural diversity of the terpenoids is related to the generation of diverse terpene skeletons by TPS [19], and these latter two enzymes are likely to regulate monoterpene and sesquiterpene distributions over the course of *A. argyi* leaf development. RT-qPCR results revealed that these six gene expression profiles were generally consistent with the FPKM values (Figure 8).

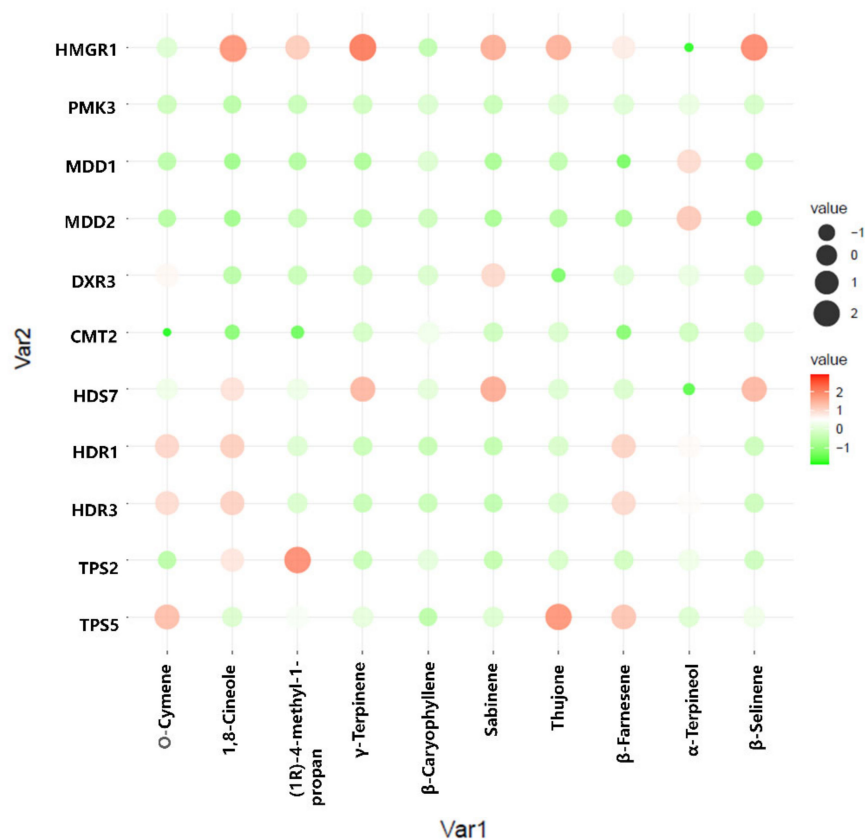


Figure 7. Pearson correlation bubble chart corresponding to gene expression patterns and terpenoid contents in *A. argyi* samples collected at four time points. The size of circles corresponds to correlation coefficient (R) values, and colors indicate whether a correlation is negative or positive.

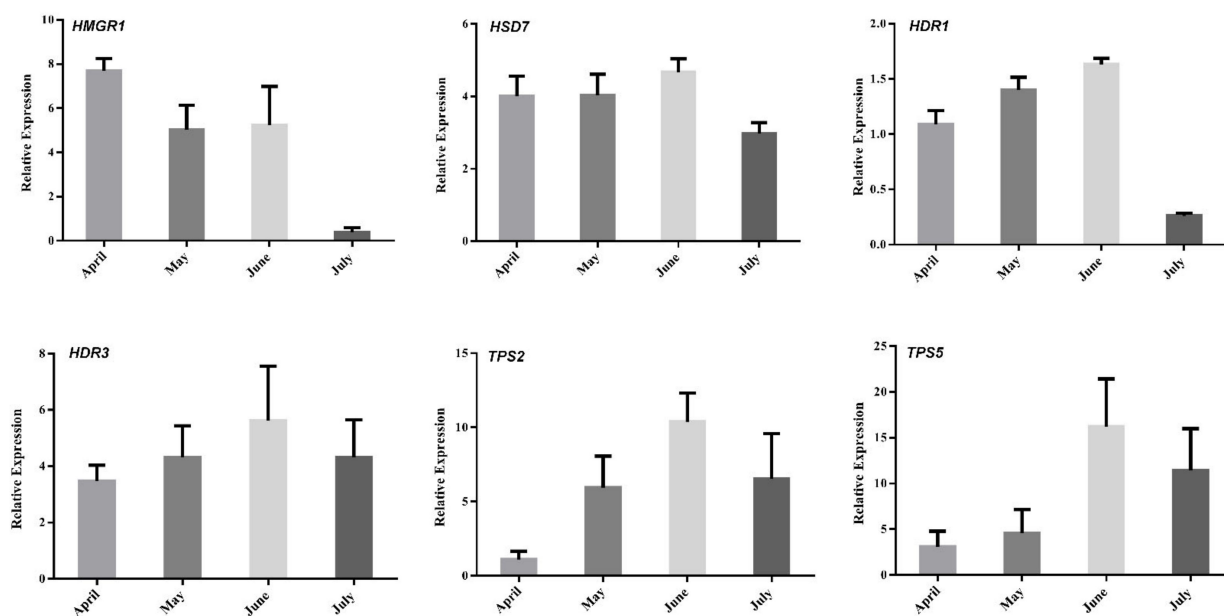


Figure 8. Expression analyses of selected genes (from Figure 7) using both qRT-PCR and RNA-Seq (FC of FPKM) approaches. Relative expression levels were estimated from the threshold of the PCR cycle, by the delta-delta Ct method. The error bars indicate the standard errors of three biological replicates.

3. Discussion

3.1. Harvest Timing Impacts the Quality and Medicinal Value of *A. Argyi* Leaves

The quality of medicinal materials is often associated with environmental parameters and other cultivation-related factors. Harvest timing, production areas, and germplasm resources are all known to be key determinants of the levels of bioactive compounds within a given plant [20]. In the present study, the quality of *A. argyi* leaves was analyzed based upon the volatile oil contents therein, revealing that these volatile oil levels (primarily composed of monoterpenes and sesquiterpenes) initially rose and then fell over the course of collection. Peak volatile oil levels were observed for samples collected in early June. This is in line with other reports, suggesting that *A. argyi* leaves harvested in June are often of higher quality [21–23]. These results also largely agree with ancient Chinese customs. Therefore, our current study reaffirmed harvest timing as the most important factor influencing the volatile oil contents within *A. argyi* leaves. Other factors not analyzed in this study, such as irrigation or fertilization strategies, can also influence the biochemical makeup of these herbal specimens and warrant further study.

3.2. PacBio Iso-Seq Sequencing and Annotation of *A. Argyi*

A. argyi is a member of the *Asteraceae* family, which is composed of over 500 species, [24] and exhibits extensive genomic complexity that has largely been unexplored to date. Most prior studies of the *A. argyi* genome were conducted based on Illumina RNA-Seq data or predictive analyses, which are well-suited for capturing coding information within mRNAs but are likely to overlook sequence information within untranslated regions (UTRs). In this study, we identified 855,813 full-length non-chimeric *A. argyi* transcript sequences via a PacBio SMRT sequencing approach, revealing that the lengths of the 5'- and 3'-UTR of most *A. argyi* mRNAs were roughly 150 and 300 nucleotides, respectively, consistent with those of *Arabidopsis*, tomato, and soybean analyzed via Sanger sequencing [25–27], confirming the high quality of such PacBio -based full-length cDNA sequencing.

Relatively few full-length transcripts are derived from Illumina RNA-Seq assemblies, and the rates of inaccurate gene structure characterization can be relatively high as a consequence of misassembly. This problem is particularly pronounced for species lacking an available reference genome [28]. SMRT Sequel is a recently developed third-generation sequencing (TGS) approach using the PacBio sequencing platform. Non-assembled long-read transcripts can be generated through this approach with a low error rate (10%), and the remaining error can be largely eliminated through RNA-seq-mediated correction [29]. Herein, we observed an *A. argyi* transcript error rate of under 1%, and, in another study, the mapping rate for maize long reads rose from 11.6% to 99.1% following Illumina correction [30]. Notably, these full-length data enabled the annotation of *A. argyi* mRNA UTRs and the construction of more accurate gene models, especially for genes expressed at lower levels and genes with long intronic regions. Overall, our data offer new insights into the *A. argyi* transcriptome, providing a valuable foundation for future studies of gene regulation in *A. argyi*.

3.3. Assessment of Differential Gene Expression in *A. Argyi*

When DEGs were identified by comparing *A. argyi* samples collected at four time points, the greatest number was observed between April and July (33,986 DEGs), while the smallest number was observed between June and July (15,351 DEGs). This suggests that many of these DEGs may play important roles in *A. argyi* growth and development in a temporal manner. Additionally, these genes were clustered into 12 clusters via a soft clustering approach, with clusters 3 and 5 being closely correlated with the volatile oil levels in *A. argyi* samples. These results were consistent with our prior data, which demonstrated the temporal patterns of unigenes expression in these plants.

A total of 1756 sequences within our dataset were associated with secondary metabolite biosynthesis, with 77 being specifically annotated as being associated with terpenoid metabolism. A majority of the carbohydrate-metabolism-related sequences were identified

in *Artemisia annua* [31]. Overall, these annotations and transcriptomic differences represent a valuable resource for the study of biosynthetic activity as a function of harvesting and developmental time in *A. argyi*.

3.4. Terpenoid Synthesis and Transport in *A. Argyi* Leaves

Prior studies have primarily employed genetic or biochemical approaches to explore the mechanisms governing terpenoid biosynthesis in *A. argyi*. Generally, terpenoid biosynthesis relies on DMAPP and IPP as precursors that are generated by the MVA and methylerythritol phosphate (MEP) pathways [7,32], with the MEP pathway being primarily involved in generating isoprenoids for the synthesis of monoterpenes [33], diterpenes [34], and carotenoids [35]. This suggests that the MEP pathway may enhance the accumulation of volatile oils by increasing the supply of precursors compounds in *A. argyi*.

HMGR is the first enzyme in the MAV pathway that plays an essential regulatory role [36,37]. In *Ginkgo biloba*, for example, the expression of this gene is correlated with terpene accumulation [13], and similar results have been observed in *Solanum tuberosum* and *Glycyrrhiza uralensis* [38,39]. The HDS and HDR genes function in the final two steps of the MEP pathway; HDS has previously been studied in *Solanum lycopersicum*, *Hevea brasiliensis*, *Arabidopsis thaliana*, and *G. biloba* [40–43], functioning as a regulator of the carbon flux of the MEP pathway and plant defense mechanisms [37], with HDR being of particular importance [44]. HDR regulates IPP and DMAPP biosynthesis in the context of andrographolide production. Herein, we conducted SMRT sequencing of *A. argyi* samples at different developmental stages and found that HDS7, HDR1, and HDR3 expression levels were positively correlated with the total volatile oil levels in analyzed leaves. As such, these genes may play an essential role in regulating the MEP pathway flux in *A. argyi*. The unigene expression levels of the MEP-pathway-related genes were higher than those for the MVA-pathway-related genes, which is consistent with the high levels of monoterpenes. Crosstalk between these two different terpene skeleton pathways has been documented, and the relative contributions of each pathway to the process of terpene biosynthesis remains uncertain [45].

The expression patterns of the TPS2 and TPS5 genes can regulate the distribution of monoterpene- and sesquiterpene-type terpenoids in the *A. argyi* samples collected during different periods. Many TPS are prolific enzymes that can produce mixtures of different proportions of the same compounds [46]. Sequence similarity alone cannot predict the specific biochemical function of a single TPS family member, because only a few amino acids can cause dramatic changes in the terpenoid structures produced by a given TPS enzyme [47,48]. Therefore, it is difficult to judge the specific functions of genes based on sequence similarity alone. The specific functions of these two genes, thus, require further verification. The future functional validation and utilization of these monoterpenoids should focus on these genes.

4. Materials and Methods

4.1. Plant Materials

A. argyi plants used in this study were planted in the botanical garden of Wuhan Polytechnic University (Wuhan, China), which is located at approximately N30.58, E114.03). Weeding was performed by hand, and micro-sprinkler systems were used for irrigation to support plant growth. No fungicides or pesticides were applied, as no relevant diseases or pests were present at the study site during the study period.

A. argyi leaves were collected monthly during the first week of April, May, June, and July from plants. Three replicate leaves were collected at each of these four time points (12 total samples), with each sample being separated in two. One half of each sample was frozen with liquid nitrogen and stored at $-80\text{ }^{\circ}\text{C}$ prior to transcriptomic analyses, while the other half was stored at $-80\text{ }^{\circ}\text{C}$ prior to measurements of the volatile oil contents.

4.2. GC–MS Analysis

4.2.1. Isolation and Concentration of Volatiles

To measure volatile oil contents, leaf samples were ground to yield a powder that was then dried to a constant weight at 25 °C in an oven. Next, 1 g (1 mL) of the powder was transferred to a 20 mL headspace vial (Agilent, Palo Alto, CA, USA), containing a NaCl-saturated solution to inhibit any enzyme reaction. The vials were sealed using crimp-top caps with TFE-silicone headspace septa (Agilent, Palo Alto, CA, USA). At the time of solid-phase micro extraction (SPME) analysis, each vial was placed at 100 °C for 5 min, then a 120 µm divinylbenzene/carboxen/polydimethylsiloxan fiber (Agilent, Palo Alto, CA, USA) was exposed to the headspace of the sample for 15 min at 100 °C.

4.2.2. GC–MS Conditions

After sampling, desorption of the VOCs from the fiber coating was carried out in the injection port of the GC apparatus (Model 8890; Agilent, Palo Alto, CA, USA) at 250 °C for 5 min in splitless mode. The identification and quantification of VOCs was carried out using an Agilent Model 8890 GC and a 5977B mass spectrometer (Agilent, Palo Alto, CA, USA), equipped with a 30 m × 0.25 mm × 0.25 µm DB-5MS (5% phenyl-polymethylsiloxane) capillary column. Helium was used as the carrier gas at a linear velocity of 1.2 mL/min. The injector temperature was maintained at 250 °C, with the detector at 280 °C. The oven temperature was programmed from 40 °C (3.5 min), increasing at 10 °C/min to 100 °C, at 7 °C/min to 180 °C, and at 25 °C/min to 280 °C, followed by a 5 min hold. Mass spectra were recorded in electron impact (EI) ionization mode at 70 eV. The quadrupole mass detector, ion source, and transfer line temperatures were set, respectively, at 150 °C, 230 °C, and 280 °C. Mass spectra were scanned in the *m/z* 50–500 amu range at 1 s intervals. Volatile compound identification was achieved by comparing the mass spectra with the data system library (Metware company-built GC–MS database (MWGC) or National Institute of Standards and Technology (NIST)) and linear retention index.

4.3. Transcriptomic Analysis

4.3.1. RNA Preparation

Samples from each plant were ground on dry ice, after which the CTAB-PBIOZOL reagent and ethanol precipitation were employed to purify total RNA from these samples based on provided instructions. RNA quality was assessed with an Agilent 2100 bioanalyzer (Thermo Fisher Scientific, Waltham, MA, USA) and an RNA 6000 Nano Labchip kit (Agilent Technologies, Santa Clara, CA, USA), and samples exhibiting an RNA integrity number between 7 and 10 were used for RNA-seq analysis, as described by Jin et al. [49].

4.3.2. PacBio Library Construction and Sequencing

Equal portions of the four total RNA samples obtained from 4 collection time points (April, May, June, and July) were pooled to produce a mixed RNA sample for PacBio sequencing. The Iso-Seq Template Preparation for Sequel Systems protocol was then employed to prepare sequencing templates. A total of 800–1000 ng of total pooled RNA was then utilized for the first-strand cDNA synthesis with the Clontech SMARTer PCR cDNA Synthesis Kit, with CDS Primer IIA first being annealed to the polyA+ tail of transcripts, after which the SMARTScribe Reverse Transcriptase was employed for first-strand synthesis. Large-scale PCR was then used to generate sufficient cDNA using Clontech PrimeSTAR GXL DNA Polymerase and 5' PCR Primer IIA (5'-AAGCAGTGGTATCAACGCAGAGTAC-3'). The DNA Damage Repair Mix (PacBio, Menlo Park, CA, USA) and End Repair Mix (PacBio) solutions were then used to repair the prepared cDNA, which subsequently underwent adapter ligation using an appropriate ligase (PacBio, Menlo Park, CA, USA), with Exo III and Exo VII being introduced into the solution to reduce unrepaired DNA or linear DNA lacking blunt adapter sequences. The remaining cDNA samples were then measured with Qubit HS (Life Technologies, Carlsbad, CA, USA) and Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA) instruments. Next, a Sequel Binding Kit 2.1 was used to

bind these prepared SMRT bell sequencing libraries to appropriate polymerases using Primers V3. A PacBio MagBead Binding Kit or diffusion loading were then used to isolate polymerase-template complexes, followed by analysis with the PacBio Sequel sequencer (BGI-Shenzhen, Shenzhen, China) using Sequel Sequencing Kit 2.1 and the Sequel SMRT Cell 1M v2 Tray. The transcriptome sequencing reads were deposited in NCBI SRA with accession number SUB10986546.

4.3.3. Illumina Library Preparation and Sequencing

After RNA from each leaf sample had been pooled in equimolar amounts, they were used for Illumina library preparation. First, mRNA was purified using oligo (dT) magnetic beads, after which it was then purified and fragmented into small pieces with a fragmentation buffer. Random hexamer primers were then used for first-strand cDNA synthesis, followed by second-strand cDNA synthesis. Samples then underwent index adapter ligation, poly-adenylation, and end repair. The resultant cDNA fragments were subjected to PCR amplification and purified with Ampure XP Beads dissolved in EB solution. An Agilent Technologies 2100 Bioanalyzer was used to ensure the quality of the resultant product, and double-stranded PCR products from the prior step were heated, denatured, and circularized with the splint oligo sequence to yield a final library, which was amplified with phi29 to generate DNA nanoballs (DNBs) containing over 300 copies of a given molecule. These DNBs were then loaded into a patterned nanoarray, and single-end 50 bp reads were obtained using a BGISEQ500 instrument (BGI-Shenzhen, Shenzhen, China). The transcriptome sequencing reads were deposited in NCBI SRA with accession number SUB10986546.

4.3.4. Full-Length Transcriptome Analysis

There were three key stages in the full-length transcriptome sequencing performed for the present study [50]: full-length sequence recognition, isoform-level clustering to ensure sequence consistency, and consistent sequence polishing. Initially, ROI sequences were filtered based on the presence of 3'- or 5'-primer sequences and poly-A tails and whether or not they contained full-length or chimeric sequences. Full-length sequences from the same isoform were then grouped using an iterative isoform-clustering algorithm, which was used to cluster the full-length sequences from the same isoform, and similar full-length sequences were clustered together. Each cluster, thus, contained a consistent sequence. Non-full-length sequences were then clustered with the Quiver algorithm, the resultant consistent sequences underwent polishing, and then high-quality sequences were screened for subsequent analysis. Only high-quality sequences were screened, as the deletion of the 5'-end of a sequence had the potential to indicate a non-full-length sequence such that only 5' exon sequences were pooled; the longest sequence was used as the final transcript sequence.

4.3.5. DEG Identification

Read counts were normalized based on calculations of the fragments per kilobase of transcript per million mapped read (FPKM) values [51], with relative expression levels further being determined based upon false discovery rate (FDR) values. Genes differentially expressed among leaves collected during four consecutive months (April, May, June, and July) of the same year were identified with the DESeq R package (1.10.1), which analyzed digital gene expression data based on a negative binomial distribution model [52]. The resultant *p*-value thresholds were corrected for multiple testing based upon FDR values [53], with an FDR < 0.01 and a fold change ≥ 2 being used to identify DEGs in this analysis.

RT-qPCR was performed on an ABI Prism 7900 Sequence Detection System (Applied Biosystems, Warrington, UK). SYBR Green Real-Time PCR Master Mix (Takara, Japan) was used for each PCR reaction in a 20 μ L reaction volume, containing 1 μ L of each primer (5 mM) and 4 μ L of first-strand cDNA. The primers used for RT-qPCR are listed in Supplementary Table S12.

4.3.6. Functional Annotation and Metabolic Pathway Analyses

Resultant mRNA sequences were compared via a BLAST approach to the NCBI non-redundant NR <ftp://ftp.ncbi.nlm.nih.gov/blast/db> (accessed on 21 October 2021), Swiss-Prot <http://ftp.ebi.ac.uk/pub/databases/swissprot> (accessed on 21 October 2021), and clusters of euKaryotic Orthologous Groups (KOG) <http://www.ncbi.nlm.nih.gov/KOG> (accessed on 21 October 2021) databases to derive NR, Swiss-Prot, and KOG annotations, respectively. Gene Ontology (GO) annotations <http://geneontology.org> (accessed on 21 October 2021) were established based on the closest BLASTX hit from the NR database using the WEGO software (E-value $\leq 10^{-5}$) [54]. KEGG <http://www.genome.jp/kegg> (accessed on 21 October 2021) pathway analyses were conducted with the KEGG Automatic Annotation Server (KAAS) at a threshold of $E \leq 10^{-5}$ [55].

5. Conclusions

In conclusion, we employed SMRT sequencing and RNA-seq strategies to efficiently identify key genes associated with the terpenoid biosynthesis in *A. argyi*. We evaluated temporal expression patterns of these metabolically important genes at four harvest time points, leading to the identification of six key genes, including one 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), one (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (HDS), two 1-deoxy-D-xylulose-5-phosphate reductoisomerase (HDR), and two terpene synthase (TPS) genes that were significantly correlated with the levels of monoterpenes or sesquiterpenes in *A. argyi* samples. Together, these data provide a robust scientific foundation for future studies exploring the molecular determinants of *A. argyi* sample quality and optimized sample harvesting, which may support genetic engineering efforts focused on the secondary metabolic pathways within relevant medicinal plants.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules27185948/s1>. Figure S1: GC–MS chromatogram in extracted ion chromatogram (EIC) mode of 10 compounds in Folium *A. argyi* extract. Figure S2: PacBio single-molecule long-read sequencing of *A. argyi*; (A) subread length distribution; (B) length distribution of full-length non-chimeric reads; (C) length distribution of CDSs of all isoforms. Figure S3: Volcano plots exhibiting the differential expression of genes associated with different *A. argyi* samples. Table S1: Identification of volatile constituents in *A. argyi* leaves collected at four time points by GC–MS. Table S2: Statistics for polymerases. Table S3: Statistics for subread results. Table S4: Statistics for CCS results. Table S5: PacBio single-molecular long-read sequencing of *A. argyi*. Table S6: Statistics for isoform results. Table S7: The number of transcripts of *A. argyi*. Table S8: Statistics for transcripts after redundant sequence removal. Table S9: Statistics for CDS prediction results. Table S10: Functional annotation and classification of assembled unigenes in *A. argyi*. Table S11: Candidate 77 transcripts involved in upstream of saponin biosynthesis and their average expression level in different harvest months (April, May, June, and July). Table S12: Primers for qRT-PCR.

Author Contributions: R.X. and Y.M. conceptualized the experiments and research plans; S.L. collected the sample; W.Z., X.C. and L.W. performed the experiments; J.G., Z.S. and S.S. carried out the data and statistical analyses; D.W. and J.Y. acquired funding; H.W. and D.W. revised the manuscript; R.X., Y.L. and C.X. wrote the article. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 81903782, 31972215 and 81872948), key technology research and demonstration project of safe and efficient production of genuine medicinal materials (No. 2020-620-000-002-04), and Foundation of Wuhan Health and Family Planning Research (No. wx16c11).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare on conflict of interest.

Abbreviations

AACT	Acetoacetyl-CoA acyltransferase
CDS	Coding sequence
CMK	4-(cytidine 50-diphospho)-2-C-methyl-D-erythritol kinase
CYP	Cytochrome
DEGs	Differentially expressed genes
DESeq	Differentially expressed sequence
DMAPP	Dimethylallyl diphosphate
DXR	1-deoxy-D-xylulose-5-phosphate reductoisomerase
FPS	Farnesyl pyrophosphate synthase
FPP	Farnesyl diphosphate
FPKM	Fragments per kilobase of exon model per million mapped reads
FPPS	Farnesyl diphosphate synthase
GC-MS	Gas chromatography–mass spectrometry
GO	Gene Ontology
HDR	(E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase
HDS	(E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase
HMGS	Hydroxy-3-methylglutaryl-CoA synthase
HMGR	3-hydroxy-3-methylglutaryl-CoA reductase;
IPP	Isopentenyl diphosphate
IPPI	Isopentenyl pyrophosphate isomerase
KEGG	Kyoto Encyclopedia of Genes and Genomes
KOG	EuKaryotic orthologous groups
MEP	2-C-methyl-D-erythritol 4phosphate
MVA	Mevalonate acid
NCBI	Nr NCBI non-redundant protein
ORF	Open read frame
ROI	Reads of insert

References

1. Jiangsu New Medical College. *Dictionary of Traditional Chinese Medicines*; Shanghai Science and Technology Publishing House: Shanghai, China, 1986; pp. 167–276.
2. Liu, T.; Liao, X.F.; WU, Y.T.; Chen, P.X.; BAI, Q.; Ren, Z.; Wang, Y.F. Research Progress on Anti-inflammatory Effective Components and Mechanism of *Artemisia argyi* Folium. *Tradit. Chin. Drug Res. Clin. Pharmacol.* **2021**, *32*, 449–453.
3. Gu, H.K.; Liu, G.J.; Cheng, Z.L. Research Advances on Application Basic Research and Utilization of *Artemisia argyi*. *J. Anhui Agric. Sci.* **2018**, *46*, 22–25.
4. Lao, A.; Fujimoto, Y.; Tatsuno, T. Studies on the constituents of *Artemisia argyi* LEVL et VANT. *Chem. Pharm. Bull.* **1984**, *32*, 723–727. [[CrossRef](#)]
5. Song, X.; Wen, X.; He, J.; Zhao, H.; Li, S.; Wang, M. Phytochemical components and biological activities of *Artemisia argyi*. *J. Funct. Foods* **2018**, *52*, 648–662. [[CrossRef](#)]
6. Jeong, M.A.; Lee, K.W.; Yoon, D.; Lee, H.J. Jaceosidin, a Pharmacologically Active Flavone Derived from *Artemisia argyi*, Inhibits Phorbol-Ester-Induced Upregulation of COX-2 and MMP-9 by Blocking Phosphorylation of ERK-1 and -2 in Cultured Human Mammary Epithelial Cells. *Ann. N. Y. Acad. Sci.* **2007**, *1095*, 458–466. [[CrossRef](#)]
7. Shu, Q.; Wang, H.; Litscher, D.; Wu, S.; Chen, L.; Gaischek, I.; Wang, L.; He, W.; Zhou, H.; Litscher, G.; et al. Acupuncture and moxibustion have different effects on fatigue by regulating the autonomic nervous system: A pilot controlled clinical trial. *Sci. Rep.* **2016**, *6*, 37846. [[CrossRef](#)]
8. Wan, L.; Lu, J.; Guo, S. GC-MS Fingerprint of Volatile Oil from *Artemisia argyi*. *Med. Plants* **2016**, *7*, 1–4.
9. Cui, L.; Wang, X.; Lu, J.; Tian, J.; Wang, L.; Qu, J. Rapid Identification of Chemical Constituents in *Artemisia argyi* Lévi. et Vant by UPLC-Q-Exactive-MS/MS. *J. Food Qual.* **2021**, *2021*, 5597327. [[CrossRef](#)]
10. Cui, Y.; Gao, X.; Wang, J.; Shang, Z.; Zhang, Z.; Zhou, Z.; Zhang, K. Full-Length Transcriptome Analysis Reveals Candidate Genes Involved in Terpenoid Biosynthesis in *Artemisia argyi*. *Front. Genet.* **2021**, *12*, 659962. [[CrossRef](#)]
11. Liu, M.; Zhu, J.; Wu, S.; Wang, C.; Guo, X.; Wu, J.; Zhou, M. De novo assembly and analysis of the *Artemisia argyi* transcriptome and identification of genes involved in terpenoid biosynthesis. *Sci. Rep.* **2018**, *8*, 5824. [[CrossRef](#)]
12. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [[CrossRef](#)] [[PubMed](#)]
13. The Gene Ontology Consortium. The Gene Ontology Consortium, Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **2017**, *45*, 331–338. [[CrossRef](#)] [[PubMed](#)]

14. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **2004**, *32*, 277–280. [[CrossRef](#)] [[PubMed](#)]
15. Koonin, E.V.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Krylov, D.M.; Makarova, K.S.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; Rao, B.S.; et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **2004**, *5*, R7. [[CrossRef](#)]
16. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2013**, *42*, D222–D230. [[CrossRef](#)]
17. Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S. UniProt: The Universal Protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, 115–119. [[CrossRef](#)]
18. Deng, Y.; Li, J.; Wu, S.; Zhu, Y.; Chen, Y.; He, F. Integrated nr database in protein annotation system and its localization. *Comput. Eng.* **2006**, *32*, 71–72.
19. Yang, M.; Liu, G.; Yamamura, Y.; Chen, F.; Fu, J. Divergent Evolution of the Diterpene Biosynthesis Pathway in Tea Plants (*Camellia sinensis*) Caused by Single Amino Acid Variation of *ent*-Kaurene Synthase. *J. Agric. Food Chem.* **2020**, *68*, 9930–9939. [[CrossRef](#)]
20. Pu, R.; Wan, D.R.; Zhao, B.X.; Wu, J.; Shi, N.N.; Huang, L.Q.; Zhang, M. Study on Effects of the Environmental Conditions on the Quality of *Artemisia argyi* Leaf in Qich. *Mod. Tradit. Chin. Med. Mater. World Sci. Technol.* **2019**, *12*, 2739–2745.
21. Hu, J.Q.; Xia, H.J.; Guo, S.X.; Zhang, R.; Li, L.Y.; Long, W.F. Study on the Detection of essential oils, total flavonoids and tannins contents in qiai and determination of optimum harvest time. *Tradit. Chin. Drug Res. Clin. Pharm.* **2016**, *31*, 3013–3016.
22. Xu, J.J.; Lu, J.Q.; Guo, S.N.; Wan, L.J. GC-MS. Analysis of Chemical Components of Essential Oil from *Artemisia argyi* in Different Parts and Different Harvest Periods. *Chin. J. Exp. Tradit. Med. Formulae* **2015**, *21*, 51–57.
23. Hong, Z.G.; Zhang, L.L.; Wu, H.G. Study on the Component and Content of Alkane from *Folium Artemisia argyi* Gathered in Different Growing Period. *J. South-Cent Univ. Natl.* **2014**, *33*, 41–43.
24. Liu, D.; Chen, Y.; Wan, X.; Shi, N.; Huang, L.; Wan, D. *Artemisiae argyi* Folium and its geo-authentic crude drug qi ai. *J. Tradit. Chin. Med Sci.* **2017**, *4*, 20–23. [[CrossRef](#)]
25. Umezawa, T.; Sakurai, T.; Totoki, Y.; Toyoda, A.; Seki, M.; Ishiwata, A.; Akiyama, K.; Kurotani, A.; Yoshida, T.; Mochida, K.; et al. Sequencing and Analysis of Approximately 40 000 Soybean cDNA Clones from a Full-Length-Enriched cDNA Library. *DNA Res.* **2008**, *15*, 333–346. [[CrossRef](#)]
26. Aoki, K.; Yano, K.; Suzuki, A.; Kawamura, S.; Sakurai, N.; Suda, K.; Kurabayashi, A.; Suzuki, T.; Tsugane, T.; Watanabe, M.; et al. Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genom.* **2010**, *11*, 210. [[CrossRef](#)]
27. Srivastava, A.K.; Lu, Y.; Zinta, G.; Lang, Z.; Zhu, J.-K. UTR-Dependent Control of Gene Expression in Plants. *Trends Plant Sci.* **2018**, *23*, 248–259. [[CrossRef](#)]
28. Li, J.; Harata-Lee, Y.; Denton, M.D.; Feng, Q.; Rathjen, J.R.; Qu, Z.; Adelson, D.L. Long read reference genome-free reconstruction of a full-length transcriptome from *Astragalus membranaceus* reveals transcript variants involved in bioactive compound biosynthesis. *Cell Discov.* **2017**, *3*, 17031. [[CrossRef](#)]
29. An, D.; Cao, H.X.; Li, C.; Humbeck, K.; Wang, W. Isoform Sequencing and State-of-Art Applications for Unravelling Complexity of Plant Transcriptomes. *Genes* **2018**, *9*, 43. [[CrossRef](#)]
30. Wang, B.; Tseng, E.; Regulski, M.; Clark, T.A.; Hon, T.; Jiao, Y.; Lu, Z.; Olson, A.; Stein, J.C.; Ware, D. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* **2016**, *7*, 11708. [[CrossRef](#)]
31. Shen, Q.; Zhang, L.; Liao, Z.; Wang, S.; Yan, T.; Shi, P.; Liu, M.; Fu, X.; Pan, Q.; Wang, Y.; et al. The Genome of *Artemisia annua* Provides Insight into the Evolution of Asteraceae Family and Artemisinin Biosynthesis. *Mol. Plant* **2018**, *11*, 776–788. [[CrossRef](#)]
32. Lange, B.M.; Rujan, T.; Martin, W.; Croteau, R. Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13172–13177. [[CrossRef](#)] [[PubMed](#)]
33. Wölwer-Rieck, U.; May, B.; Lankes, C.; Wüst, M. Methylerythritol and Mevalonate Pathway Contributions to Biosynthesis of Mono-, Sesqui-, and Diterpenes in Glandular Trichomes and Leaves of *Stevia rebaudiana* Bertoni. *J. Agric. Food Chem.* **2014**, *62*, 2428–2435. [[CrossRef](#)] [[PubMed](#)]
34. Ge, X.; Wu, J.-Y. Induction and potentiation of diterpenoid tanshinone accumulation in *Salvia miltiorrhiza* hairy roots by β -aminobutyric acid. *Appl. Microbiol. Biotechnol.* **2005**, *68*, 183–188. [[CrossRef](#)] [[PubMed](#)]
35. Lu, X.-M.; Hu, X.-J.; Zhao, Y.-Z.; Song, W.-B.; Zhang, M.; Chen, Z.-L.; Chen, W.; Dong, Y.-B.; Wang, Z.-H.; Lai, J.-S. Map-Based Cloning of *zb7* Encoding an IPP and DMAPP Synthase in the MEP Pathway of Maize. *Mol. Plant* **2012**, *5*, 1100–1112. [[CrossRef](#)]
36. Cordoba, E.; Salmi, M.; León, P. Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.* **2009**, *60*, 2933–2943. [[CrossRef](#)]
37. Gil, M.J.; Coego, A.; Mauch-Mani, B.; Jordá, L.; Vera, P. The Arabidopsis *csb3* mutant reveals a regulatory link between salicylic acid-mediated disease resistance and the methyl-erythritol 4-phosphate pathway. *Plant J.* **2005**, *44*, 155–166. [[CrossRef](#)]
38. Chen, D.H.; Ye, H.C.; Li, G.F.; Liu, Y. Cloning and sequencing of HMGR gene of *Solanum tuberosum* and its expression pattern. *Acta Bot. Sin.* **2000**, *42*, 724–727.
39. Liu, Y.; Xu, Q.-X.; Wang, X.-Y.; Liu, C.-S.; Chen, H.-H. Analysis on correlation between 3-hydroxy-3-methylglutaryl-coenzyme A reductase gene polymorphism of *Glycyrrhiza uralensis* and content of glycyrrhizic acid. *China J. Chin. Mater. Medica* **2012**, *37*, 3789–3792.

40. Querol, J.; Imperial, S.; Boronat, A. Bioinformatic and molecular analysis of hydroxymethylbutenyl diphosphate synthase (GCPE) gene expression during carotenoid accumulation in ripening tomato fruit. *Planta* **2003**, *217*, 476–482. [[CrossRef](#)]
41. Sando, T.; Takeno, S.; Watanabe, N.; Okumoto, H.; Kuzuyama, T.; Yamashita, A.; Hattori, M.; Ogasawara, N.; Fukusaki, E.; Kobayashi, A. Cloning and Characterization of the 2-C-Methyl-D-erythritol 4-Phosphate (MEP) Pathway Genes of a Natural-Rubber Producing Plant, *Hevea brasiliensis*. *Biosci. Biotechnol. Biochem.* **2008**, *72*, 2903–2917. [[CrossRef](#)]
42. Gutiérrez-Nava, M.D.L.L.; Gillmor, C.S.; Jiménez, L.F.; Guevara-García, A.; León, P. *CHLOROPLAST BIOGENESIS* Genes Act Cell and Noncell Autonomously in Early Chloroplast Development. *Plant Physiol.* **2004**, *135*, 471–482. [[CrossRef](#)]
43. Sang, M.K.; Soo-Un, K. Characterization of 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase (HDS) gene from *Ginkgo biloba*. *Mol. Biol. Rep.* **2009**, *37*, 973–979.
44. Laule, O.; Furholz, A.; Chang, H.-S.; Zhu, T.; Wang, X.; Heifetz, P.B.; Gruissem, W.; Lange, M. Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6866–6871. [[CrossRef](#)]
45. Qiu, F.; Wang, X.; Zheng, Y.; Wang, H.; Liu, X.; Su, X. Full-Length Transcriptome Sequencing and Different Chemotype Expression Profile Analysis of Genes Related to Monoterpenoid Biosynthesis in *Cinnamomum porrectum*. *Int. J. Mol. Sci.* **2019**, *20*, 6230. [[CrossRef](#)]
46. Martin, D.M.; Fäldt, J.; Bohlmann, J. Functional Characterization of Nine Norway Spruce *TPS* Genes and Evolution of Gymnosperm Terpene Synthases of the *TPS-d* Subfamily. *Plant Physiol.* **2004**, *135*, 1908–1927. [[CrossRef](#)]
47. Zulak, K.G.; Lippert, D.N.; Kuzyk, M.A.; Domanski, D.; Chou, T.; Borchers, C.H.; Bohlmann, J. Targeted proteomics using selected reaction monitoring reveals the induction of specific terpene synthases in a multi-level study of methyl jasmonate-treated Norway spruce (*Picea abies*). *Plant J.* **2009**, *60*, 1015–1030. [[CrossRef](#)]
48. Dudareva, N. (E)-beta-Ocimene and Myrcene synthase genes of floral scent biosynthesis in snapdragon: Function and expression of three terpene synthase genes of a new terpene synthase subfamily. *Plant Cell* **2003**, *15*, 1227–1241. [[CrossRef](#)]
49. Jin, J.; Panicker, D.; Wang, Q.; Kim, M.J.; Liu, J.; Yin, J.-L.; Wong, L.; Jang, I.-C.; Chua, N.-H.; Sarojam, R. Next generation sequencing unravels the biosynthetic ability of Spearmint (*Mentha spicata*) peltate glandular trichomes through comparative transcriptomics. *BMC Plant Biol.* **2014**, *14*, 292. [[CrossRef](#)]
50. Ye, J.; Cheng, S.; Zhou, X.; Chen, Z.; Kim, S.U.; Tan, J.; Zheng, J.; Xu, F.; Zhang, W.; Liao, Y.; et al. A global survey of full-length transcriptome of *Ginkgo biloba* reveals transcript variants involved in flavonoid biosynthesis. *Ind. Crop. Prod.* **2019**, *139*, 111547. [[CrossRef](#)]
51. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R. Differential gene and transcript expression analysis of RNA-seq experiments with *TopHat* and *Cuffiinks*. *Nat. Protoc.* **2012**, *7*, 562–578. [[CrossRef](#)]
52. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* **2010**, *26*, 139–140. [[CrossRef](#)]
53. Orr, M.; Liu, P.; Nettleton, D. An improved method for computing q-values when the distribution of effect sizes is asymmetric. *Bioinformatics* **2014**, *30*, 3044–3053. [[CrossRef](#)]
54. Ye, J.; Fang, L.; Zheng, H.; Zhang, Y.; Chen, J.; Zhang, Z.; Wang, J.; Li, S.; Li, R.; Bolund, L.; et al. WEGO: A web tool for plotting GO annotations. *Nucleic Acids Res.* **2006**, *34*, W293–W297. [[CrossRef](#)]
55. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **2007**, *35* (Suppl. 2), W182–W185. [[CrossRef](#)]