Article

Identification of Key Genes for Oleoresin Biosynthesis in High and Low Oleoresin-Yielding Slash Pine Based on Transcriptome Analysis

Min Yi [1], **Lu Zhang** [1], **Zishan Cheng** [1], **Rong Hu** [1], **Yuan Gao** [1], **Cangfu Jin** [1], **Shenggui Yuan** [1], **Shiwu Sun** [2] **and Meng Lai** [1,*]

1 Key Laboratory of Silviculture, College of Forestry, Jiangxi Agricultural University, Nanchang 330045, China
2 Baiyun Mountain State Owned Forest Farm, Jian 343011, China
* Correspondence: laimeng21@163.com

Abstract: Slash pine (*Pinus elliottii* Engelmann) is a pine species widely cultivated for its high oleoresin production capacity. However, little is known about the underlying molecular mechanism of oleoresin biosynthesis between high and low oleoresin-yielding slash pines. In this study, the terpenoid compositions of oleoresin harvested from high- and low-yielding slash pines were identified using gas chromatography/mass spectrometry (GC-MS) analysis. The monoterpenes and diterpenes are the major constituents, of which the α- and β-pinenes are the overwhelming majority of turpentines, and abietic acid, levopimaric acid, and neoabietic acid are the most abundant in rosin. The transcriptomic analysis was also performed with secondary xylem tissues of high- and low-yielding slash pines. After functional annotation, the DEGs of RNA-seq data between high- and low-yielding pines in April, July, and October were screened, and many key enzyme genes were found to be implicated in terpenoid backbone biosynthesis. Moreover, weighted gene correlation network analysis (WGCNA) was carried out to uncover the gene modules highly related to α- and β-pinene biosynthesis in slash pine. Twenty-three modules were attained in this study. Focusing on the total oleoresin yield, the MEblue module exhibited the highest positive correlation, while the MEgreen module exhibited the highest negative correlation. A total of 20 TFs were identified in gene modules. Among these genes, the *c215396.graph_c0* encoding an MYB TF is the key differentially expressed gene (DEG) between high- and low-yielding pines. The subsequent one-hybrid yeast assay verified that *c215396.graph_c0* can activate the transcription of *Apetala 2* (*AP2*) and *1-deoxy-D-xylulose 5-phosphate synthase* (*dxs*), which are also two DEGs between high- and low-yielding pines. Thus, our study identified a set of key enzymes and TFs that are involved in regulating oleoresin and composition between high- and low-yielding slash pines and provided us a deep insight into oleoresin biosynthesis.

Keywords: *Pinus elliottii* Engelmann; comparative transcriptomic analysis; resin yield; oleoresin biosynthesis; WGCNA

1. Introduction

Pine oleoresin, derived from the Pinus species, is a complex viscous mixture of terpenes that includes two major portions: the volatile turpentine and the nonvolatile rosin. Turpentine consists of mono-(C_{10}) and sesquiterpene (C_{15}), while diterpene (C_{20}) is the major component of rosin. In conifers, terpenes primarily function as the insect and microbial toxins to defend plants against the invasion of aggressive pests or pathogenic fungi [1–3]. Moreover, monoterpenes emitted by vegetation exhibit antioxidant properties and are implicated in the oxidative stress accommodation and thermotolerance of plants [4,5]. Nowadays, pine oleoresin is one of the most important nonwood forestry products, of which the derivatives are widely used in numerous industrial productions, such as daily chemicals, building materials, food additives, pharmaceuticals, and insecticides, among others [6].

As the foremost source of terpenoids, natural vegetation provides over 30,000 terpenes with enormous structural diversities and complexities [1,7,8]. The biosynthesis of terpenes initiates with the formation of two substrates, the simple C_5-unit isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). Two separate metabolic pathways, the mevalonate (MVA) and methyl-erythritol 4-phosphate (MEP) pathways, are responsible for the formation of C_5-unit IPP and DMAPP in cytosol and plastids, respectively [9]. The synthesized DMAPP and IPP are the building blocks that undergo further condensation reactions to form the direct precursors of terpenes, including the C_{10}-unit liner prenyl diphosphpates geranyl diphosphate (GPP), C15-unit farnesyl diphosphate (FPP), and C_{20}-unit geranylgeranyl diphosphate (GGPP) [10]. These condensation reactions are catalyzed by the short-chain isoprenyl diphosphate synthases (IDSs) that belong to a large enzyme class, also known as prenytransferases [11,12]. In the MEP pathway, geranyl diphosphate synthase (GPP synthase) condenses one molecular DMAPP with two IPP molecules to produce GPP, while the formation of GGPP needs one molecular DMAPP and three IPP molecules under the catalyzation of geranylgeranyl diphosphate synthase (GGPP synthase) in plastids [10,13,14]. In contrast, the biosynthesis of FPP from DMAPP and one molecular of IPP is primarily mediated by farnesyl pyrophosphate synthases (FPP synthase) that function downstream of the MVA pathway occurring in cytosol [10,14]. As the final step, these short-chain prenyl diphosphate intermediates, such as GPP, FPP and GGPP, are catalyzed by specific terpene synthases (TPS) to synthesize diverse terpenes [10]. Sesquiterpene synthase catalyzes FPP to form sesquiterpenes, while monoterpene and diterpene synthases catalyze the formation of monoterpenes and diterpenes from GPP and GGPP, respectively [10]. After the formation of these basic terpene units, monoterpenes, diterpenes, and sesquiterpenes can be subject to further modifications, including reduction, fusion, hydroxylation, glycosylation, cyclization, and prenylation reactions to transform into other metabolites [1,10,15,16]. Many key genes in the MEP pathway are highly associated with oleoresin production, including *4-hydroxy-3-methylbut-2-enyl diphosphate reductase* (*HDR*), *DXS*, and *geranylgeranyl pyrophosphate synthase*(*GGPS*) [17]. Moreover, (-)-α/β-pinene synthase and ethylene-responsive transcription factors (ERFs) are implicated in regulating the yield of oleoresin in Masson pine [18,19]. Additionally, previous studies demonstrated that the biosynthesis of terpenoid in conifers is also affected by diverse biotic and abiotic stresses, such as senescence, wounding, hormones, pathogens, and pests [20].

Slash pine is an important timber resource and resin-producing species are native to southeastern United States. With the characterization of fast growth, strong adaptability, long fiber, and high yield of secondary metabolites, such as oleoresin via the ducts, slash pine is widely planted in many countries for its economic value [21–24]. In Brazil and China, slash pine is the major conifer species for the production of pine resin [25]. Generally, the oleoresin of slash pine is annually tapped through the bark-chipping method. Although it is labor intensive, this traditional manual harvesting way is still the common method to harvest resin [26]. Over the last two decades, numerous studies were conducted to explore the possibilities in increasing the resin yield from slash pine, such as developing new resin tapping methods [26,27], exploring efficient resin stimulant paste during resinosis [28], as well as screening for high resin-yielding pine individuals [29]. However, the lack of knowledge about the genetic regulation network of oleoresin biosynthesis in slash pine severely impedes us in developing new breeding method.

In the present study, the overall aim was to explore the terpenoid composition of oleoresin produced by high- or low-yielding slash pines, and through transcriptomic analysis and WCGNA analysis to identify the key genes and gene modules in relation to the terpenoid biosynthesis of high and low resin-yielding slash pines. The ultimate goal was to establish a gene regulation network controlling the terpenoid biosynthesis of slash pine.

2. Materials and Methods

2.1. Plant Materials and Tissue Sampling for RNA-Seq

In 2016, resin yields (RY) of 1257 healthy slash pines located in BaiYun Shan Forest, Jian city, Jiangxi Province, southern China (27.22° N, 115.13° E) were measured once a month from April to October by bark streak wounding method [25]. Among them, 16 individuals of high RY (yielding > 10.43 kg/year) and 16 individuals with low RY (yielding < 2.51 kg/year) were selected. The samples were segregated into two groups (high and low) and four individuals were pooled, producing four biological repeats for each group. Moreover, their secondary xylem tissues were collected on April, July, and October. For sampling, a square of 10 cm long by 12 cm wide from the trunk approximately 1 m–1.3 m above the ground was selected, and the bark, phloem, and cambium were sequentially removed. Finally, twenty-four 5 mm deep secondary xylem tissues were harvested. The samples were immediately placed in liquid nitrogen and then stored at −80 °C for RNA-seq.

2.2. GC-MS Analysis of Oleoresin Compositions

The fresh oleoresin collected from slash pines was separately transferred into microcentrifuge tubes and sealed with Parafilm (USA). Subsequently, the sealed tubes were placed in dry ice and stored at −80 °C. Finally, the terpenoid compositions of each sample were identified by GC-MS analysis, according to the method by Lai [30].

2.3. cDNA Library Construction

The RNA samples were assembled by 1 μg RNA input from each specimen. With the suggestion of the manufacturer, sequencing libraries were initiated by using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, Ipswich, MA, USA). The poly-T oligo-attached magnetic beads were employed to extract the mRNA from total RNA. The NEB-Next First Strand Synthesis Reaction Buffer (5×) was used to bring out fragmentations. The synthesis of the first strand cDNA was performed by using random hexamer primer and M-MuLV Reverse Transcriptase, while the second strand cDNA was subsequently synthesized by using DNA Polymerase I and RNase H. The desired cDNA pieces of 240 bp in length were purposefully harvested using AMPure XP system (Beckman Coulter, Beverly, Brea, CA, USA) from the library bits. Subsequently, the standardized cDNA specimens were mixed with 3 μL USER Enzyme (NEB, Ipswich, MA, USA) under 37 °C for 15 min and then 95 °C for 5 min. Thereafter, PCR was completed by using the Index (×) Primer, Phusion High-Fidelity DNA polymerase, and Universal PCR primers. Finally, the PCR products were purified (AMPure XP system) and the libraries were analyzed using the Agilent Bioanalyzer 2100 system.

2.4. Clustering and Sequencing

The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumia, Bologna, Italy) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Hiseq 2000 platform and paired-end reads were generated.

2.5. Sequence Data Analysis and Annotation

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N and low-quality reads from raw data. At the same time, Q20, Q30, GC-content, and sequence duplication level of the clean data were calculated. All the downstream analyses were based on clean data of high quality. The left files (read1 files) from all libraries/samples were pooled into one big left.fq file, and right files (read2 files) into one big right.fq file. Transcriptome assembly was accomplished based on the left.fq and right.fq using Trinity [31] with min_kmer_cov set to 2 by default and all other parameters set to default. Gene function was annotated based on the following databases: NCBI

non-redundant protein sequences (NR), Protein family (Pfam), Clusters of Orthologous Groups of proteins (KOG/COG/eggNOG), A manually annotated and reviewed protein sequence database (Swiss-Prot), Kyoto Encyclopedia of Genes and Genomes(KEGG), and Gene Ontology (GO).

2.6. Differential Expression Analysis

The expression abundance of the corresponding Unigenes were represented by the fragments per kilobase of transcript per million mapped reads (FPKM) value. The DESeq2 R package (1.16.1) was utilized to perform differential expression analysis targeting 6 groups with four biological replicates per condition. This study compared the specific expression genes of the high- and low-yielding varieties in April (P4), July (P7), and October (P10). DESeq2 used a model based on the negative binomial dispersal to deliver a statistical routine for the resolution of differential expression in digital gene expression data. False discovery rate was prevented by the derived p-values refined with the Benjamini and Hochberg's approach. The adapted p-values ≤ 0.05 manufactured by DESeq2 screened the differentially expressed genes. The DEGs were adopted according to the FDR ≤ 0.001 together with the absolute value of Log2 (relative expression level) ≥ 1. For clustering, we used hierarchical clustering method, and heat maps were produced from the "heatmap" function in R.

2.7. GO and KEGG Enrichment ANALYSIS

GO enrichment analysis of the DEGs was implemented by the topGO R packages based on Kolmogorov–Smirnov test. KOBAS [32] software was used to test the statistical enrichment of DEGs in KEGG pathways. KEGG pathways with a corrected p value < 0.05 were considered statistically significant.

2.8. Plant RNA Extraction and Real-Time RT-PCR

The Trizol reagent (TAKARA, Kyoto, Japan) was used for extraction of the total RNA. Then, 500 ng of total RNA was employed for the reverse transcription reactions using the Transcriptor First Strand cDNA Synthesis Kit (TAKARA, Kyoto, Japan). The cDNA was diminished 10 times, then deployed as a template for the real-time PCR (RT-PCR). The RT-PCR reaction system (Thermofisher 7500) was composed of the PCR Forward Primer (10 μM) 0.2 μL, SYBR® Premix Ex TaqTM (2×) 5 μL, cDNA template 1 μL, PCR Reverse Primer (10 μM) 0.2 μL accompanied by distilled and deionized water, up to 10 μL. Primers used for the RT-PCR are listed in the Supplementary Materials Table S1.

2.9. WGCNA Analysis

The OmicShare tools (accessed on 20 November 2020, www.omicshare.com/tools) were used to generate the hierarchical cluster regarding the PC genes (mean FPKM ≥ 1). The transcript expression matrix [33] provided the foundation for constructing the unintended co-expression networks with the aid of WGCNA (v1.47). The network formation and module detection were achieved through the application of 'cutreeDynamic' and 'mergeCloseModules' with settings power as 14, the minModuleSize as 30, and the cutHeight as 0.25. The relationships between the gene modules and samples were examined. The result revealed a significant correlation between modules and the subject of α-pinene, β-pinene, myrcene, diterpene, camphene, isopimaric acid, abietic acid, levopimaric acid, etc. The genes in the modules were further categorized through GO enrichment analysis.

2.10. Yeast One-Hybrid Assay

The genomic DNA of vascular cambium material was extracted using CTAB method [34]. The binding capacity of the MYB TF (*c215396.graph_c0*) to the 5′ UTR of *c183227.graph_c0* and *c193143.graph_c1* was assessed through one-hybrid system in the YM4271 yeast strain, and 5′ UTR fragments were exploited as correspondents to be constructed into pLacZi plasmids [35]. The gene sequence of *c215396.graph_c0* was constructed in pGADT7-AD

plasmids. The sequences are listed in the Supplementary Materials Data S1, and the primers used are listed in the Supplementary Materials Table S1.

3. Results

3.1. Compositions of Oleoresin Compounds in High- and Low-Yielding Slash Pine

In order to analyze the terpenoid composition of the oleoresin produced by slash pine, GC-MS analysis was carried out. The oleoresin of 16 slash pines with the highest oleoresin yields (yields > 10.43 kg/year) and of 16 slash pines with the lowest oleoresin yields (yields < 2.51 kg/year) were analyzed. In total, 20 oleoresin terpenoids with the content higher than 0.1% were identified in the oleoresin of both high- and low-yielding slash pines, including 5 monoterpenes and 15 diterpenes (Supplementary Materials Table S2). The percentages of 20 oleoresin terpenoids in total oleoresin of high- or low-yielding slash pine harvested in April, July, and October are listed in Supplementary Materials Table S2. The α- and β-pinenes are the most abundant terpenoids, which account for at least 40% of total oleoresin from high- or low-yielding slash pine. In regarding to diterpenes, abietic acid, levopimaric acid, and neoabietic acid are the most abundant terpenoids in oleoresin from high- or low-yielding slash pine.

In the oleoresin of high- and low-yielding slash pine, diterpenes account for more than 50%, while monoterpenes occupy a lower proportion than diterpenes. Additionally, the low-yielding slash pine has more α-pinene in the percentage of total oleoresin than high-yielding slash pine.

3.2. RNA-Seq Analysis

Xylem tissues of the four highest and four lowest oleoresin-yielding slash pines from three different oleoresin-secreting stages (April, July, and October) were sampled (Figure 1A), of which the total RNA was isolated and further proceeded to undergo direct RNA-sequencing. After data qualifying and screening, a total of 196.47 Gb of clean data were generated from the RNA-seq data of 24 samples. The average GC contents of sequencing data from 24 libraries were approximately 45%, and the Q30 value of all samples was >93%. Moreover, the mapping ratio of each sample ranged from 68.59% to 81.53%. Following this, a total number of 91,140 transcripts and 42,413 unigenes were assembled from the short-read sequences by using the Trinity assembly program. The total length of transcript is 142,747,613 bp, and the unigene has a total length of 60,728,789 bp. Additionally, N50 lengths of the transcripts and unigenes were 2217 bp and 2138 bp, respectively. Of all 42,413 unigenes, 20,935 unigenes (49.36%) are shorter than 1000 nt, whereas 21,478 unigenes (50.64%) are longer than 1000 nt. The total unigenes have a mean length of 1431.84 nt. In summary, these results indicated that the RNA-seq yielded a high quality data that showed high assembly integrity. Additionally, the principal component analysis (PCA) and the correlation analysis suggested the high reproducibility of the RNA-seq data (Figure 1B).
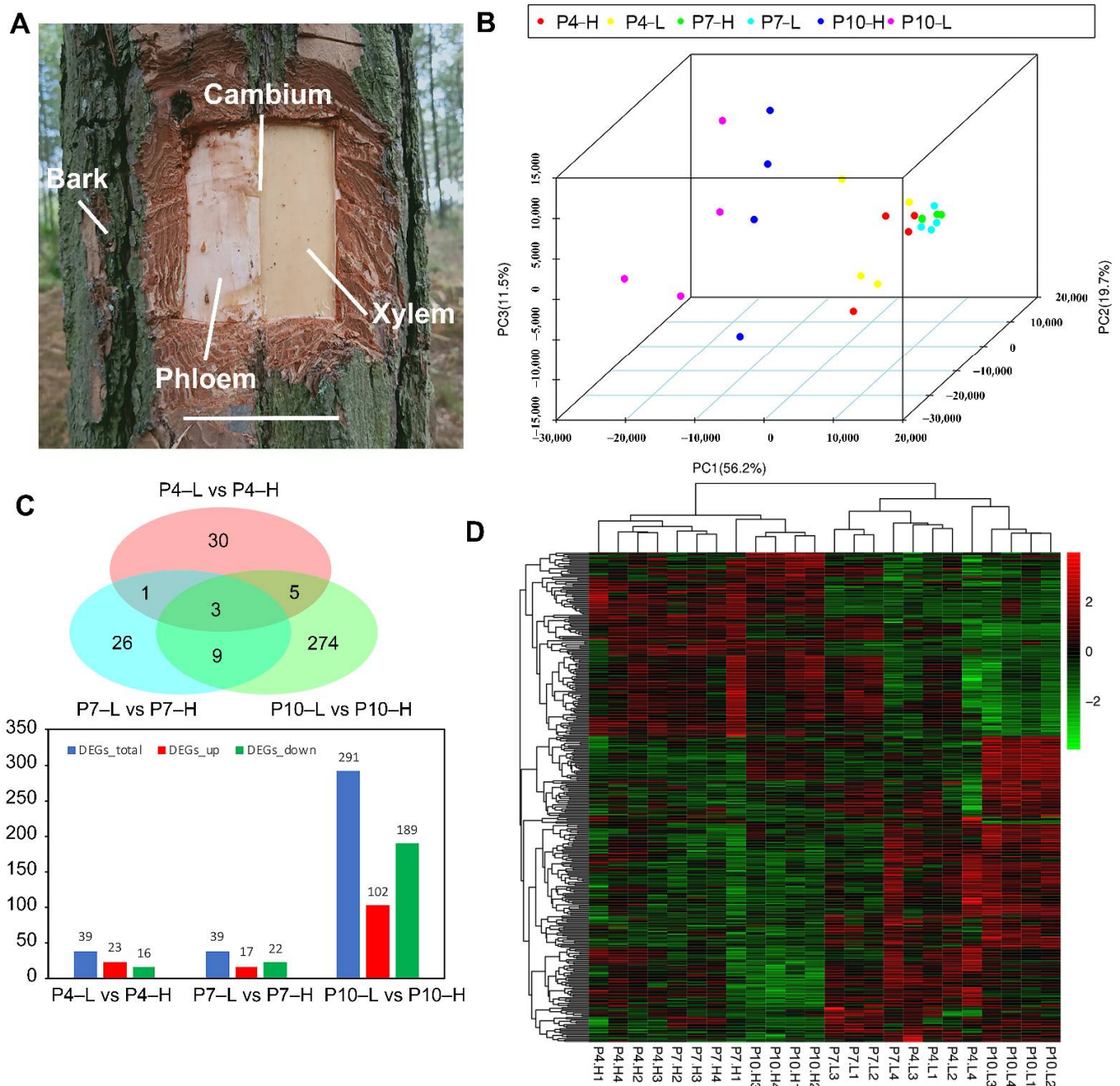
Figure 1. The overall DEGs between the high- and low-yielding slash pines. (**A**) Detailed sampling sites of high- and low-yielding slash pines for RNA-seq. White bar = 10 cm. (**B**) Principal component analysis (PCA) of 24 samples based on the DEGs (Log2 $\geq$ 1 or $\leq$ −1). (**C**) Venn diagram and histogram. The venn diagram shows the number and distribution of DEGs between high- and low-yielding slash pines in April, July, and October, and the histogram indicates DEGs numbers. The blue stands for the total DEGs, red means the up-expressed DEGs, and green is the repressed DEGs. (**D**) Heat map exhibits the congregation of relative expression levels in different groups.

3.3. Functional Annotation

Functional annotation of the assembled unigenes was performed to classify their potential functions in slash pine. To this end, sequence alignment based on similarity was carried out with the several public databases, including NR, Swiss-Prot, GO, COG, KOG, KEGG, and Pfam using BLAST software with E-value $< 1 \times 10^{-5}$. As a result, a total of 30,330 unigenes were annotated, with 9782; 16,225; 11,170; 17,782; 22,178; 18,980; 27,598; and 27,405 genes distributed in the NR, Swiss-Prot, GO, COG, KOG, KEGG, and Pfam

databases, respectively. In addition, 11,417 unigenes were identified with 300 nt ≤ lengths ≤ 1000 nt, and 18,913 unigenes exhibited the length ≥1000 nt.

To investigate the DEGs in relation to oleoresin yield of slash pine, the RNA-seq data of 24 samples were classified into six sequencing libraries based on high- or low-yielding slash pine. The changes in the transcriptome between high- and low-yielding slash pines at different oleoresin-yielding stages (April, July, and October) were analyzed. Three comparisons of P4-L vs. P4-H, P7-L vs. P7-H, and P10-L vs. P10-H (H or L refer to high or low oleoresin-yielding slash pine) were conducted (Supplementary Materials Tables S3–S5). In P4-L vs. P4-H, a total of 39 DEGs were identified with 23 upregulated and 16 downregulated genes. In P7-L vs. P7-H, there were 39 DEGs identified with 17 upregulated and 22 downregulated genes. Significantly, 291 DEGs were identified between P10-L and P10-H, with 102 upregulated and 189 downregulated genes. In Figure 1C, the Venn diagram shows that there are 30, 26, and 274 unique DEGs in P4-L vs. P4-H, P7-L vs. P7-H, and P10-L vs. P10-H groups, respectively. Only three DEGs are commonly expressed across all three compared groups. The list of all DEGs of P4-L vs. P4-H, P7-L vs. P7-H, and P10-L vs. P10-H is presented in Supplementary Materials Table S4.

The functions of these DEGs were further annotated by aligning their sequences with public databases. According to GO enrichment analysis, all of the DEGs between high and low oleoresin-yielding slash pines were enriched to three classes, including biological process, cellular components, and molecular function. In the biological process, the significant two enriched GO terms are "metabolic process" and "cellular process". In cell component, the primary term is "cell and cell part", and the largest proportion in molecular function is "catalytic activity and binding".

In KEGG enrichment analysis, a total of 20, 9, and 111 DEGs in P4-L vs. P4-H, P7-L vs. P7-H, and P10-L vs. P10-H were annotated, respectively. The DEGs of P4-L vs. P4-H and P7-L vs. P7-H were significantly enriched in the "galactose metabolism" pathway. In regard to the terpenoid biosynthesis (Supplementary Materials Table S6), DEGs of P4-L vs. P4-H were also enriched in "diterpenoid biosynthesis" and "carotenoid biosynthesis" pathways, while DEGs of P10-L vs. P10-H were significantly enriched in "ubiquinone and other terpenoid biosynthesis" pathway. However, DEGs of P7-L vs. P7-H were not enriched in any known terpenoid biosynthesis-related pathways.

3.4. DEGs Related to Terpenoids Biosynthesis

In the present study, there were a total of 34 DEGs between high and low oleoresin-yielding slash pines identified to function in the MEP and MVA pathways that are associated with the biosynthesis of terpenoids (Figure 2A). Annotated DEGs associated with 1-deoxy-D-xylulose 5-phosphate synthases (*dxs*), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (*ispD*), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthases (*ispF*), and (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (*ispG*) in the MEP pathway, and mevalonate kinase, phosphomevalonate kinase, as well as diphosphomevalonate decarboxylase in the MVA pathway encode enzymes that are involved in the biosynthesis of IPP and its isomer DMAPP. In addition, there were also DEGs encoding (2Z,6Z)-farnesyl diphosphate synthase (*ZFPS*) (*c213774.graph_c0*), geranylgeranyl diphosphate synthases (*idsA*) (*c228743.graph_c0* and *c216266.graph_c1*), and all-trans-nonaprenyl-diphosphate synthase (*SPS*) (*c206960.graph_c0*) that catalyze the condensation reactions to form GPP, FPP, and GGPP. There were also nine DEGs associated with TPS, including six isoprene synthases (*ispS*), two (+)-α-pinene synthases (*PT30*), and one (−)-β-pinene synthase(*QH6*) encoding genes(Figure 2A). Intriguingly, not all these abovementioned DEGs were upregulated in high-yield slash pine. For instance, *QH6* (*C204351.graph_c1*), *idsA* (*c216266.graph_c1*), and *ZFPS* (*c213774.graph_c0*) were significantly downregulated in high-yielding slash pine at different stages. Moreover, the DEGs in relation to MEP pathway had dramatically higher expression levels in high-yielding slash pine than in low-yielding slash pine, whereas this penomenon was not observed in the

expression profile of DEGs related to the MVP pathway. These results indicate that the high oleoresin yield of high-yielding slash pine is attributed to the enhanced terpenoid biosynthesis mediated by the MEP pathway.
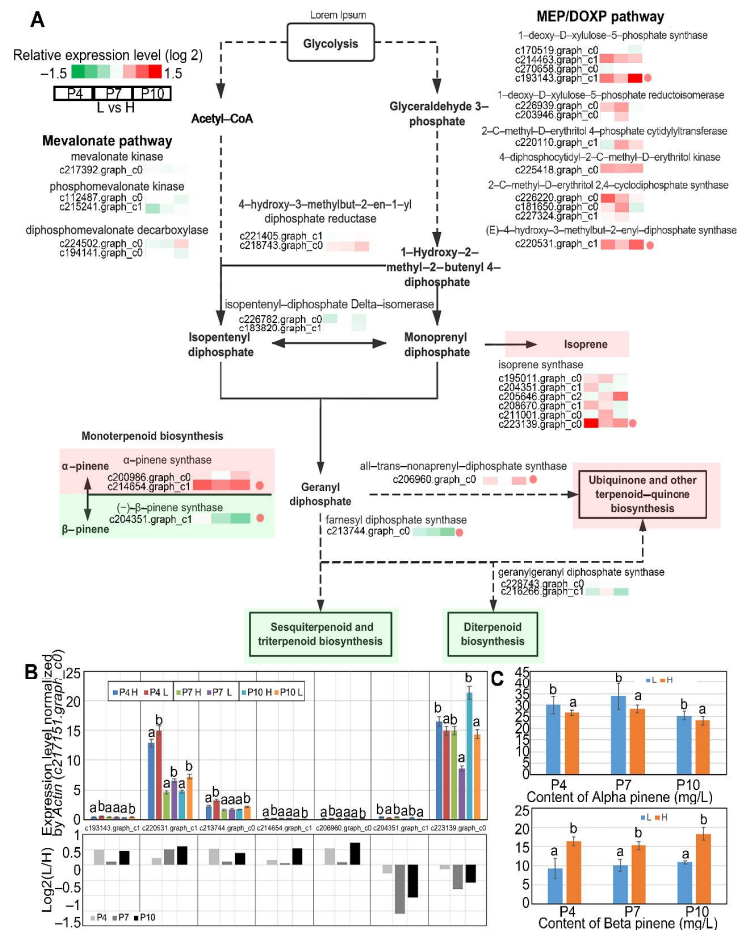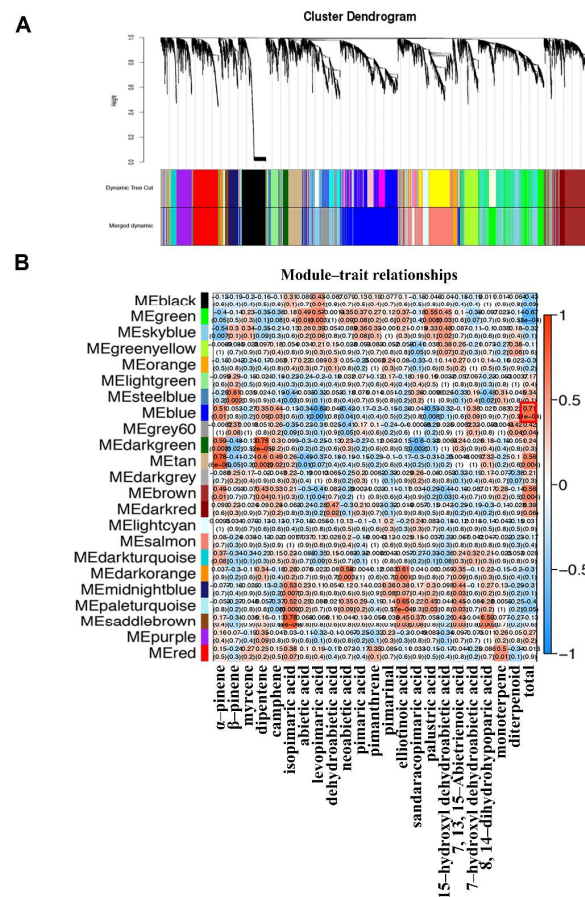

Figure 2. DEGs associated with the biosynthesis of terpenoids and steroids. (**A**) Typical DEGs in the biosynthesis of terpenoids and steroids pathway. The boxes colored red and green represent upregulated and downregulated genes, respectively. H or L refer to high or low oleoresin-yielding slash pine. (**B**) Results of RT-PCR of 7 selected DEGs between high- and low-yielding slash. All exhibitions were standardized by the expression level of *Actin* (*c2171751.graph_c0*). (**C**) The content of α- and β-pinenes in the vascular cambium of high- and low-yielding slash pines at P4, P7, and P10. All bars represent means ± s.d and those columns in histograms labeled with different letters are significantly different with *p* value < 0.05 by Duncan's test. Four biological replications in the experiment were carried out.

To confirm the reliability of the results of DEGs from RNA-seq, seven representative DEGs were selected for qRT-PCR analysis (Figure 2B). The relative expression levels of *c204351.graph_c1* and *c223139.graph_c0* in the low-yielding slash pine were lower than high-yielding slash pine, while the other five genes, *c214654.graph_c1*, *c193143.graph_c1*, *c220531.graph_c1*, *c206960.graph_c0*, and *c213744.graph_c0*, were highly expressed in the high-yielding slash pine. Moreover, the percentage proportion of α-pinene in the oleoresin produced by the high-yielding slash pine is remarkably lower than low-yielding slash pines, while the β-pinene found in the high-yielding slash pine is rather notably higher than in low-yielding slash pines (Figure 2C). Collectively, the data of the qRT-PCR and pinene content measuring results were consistent with the RNA-seq results, indicating the high level of reliability of the RNA-seq data used in this study.

### 3.5. Co-Expressed Modules Related to the Total Oleoresin Biosynthesis

The WGCNA is an efficient bioinformatics method that can be used to describe the correlation patterns among genes across microarray data, to find gene modules with similar expression pattern, and to identify hub genes in connecting gene modules to one another [33]. A total of 30,330 unigenes were selected from the RNA-seq datasets of high and low oleoresin-yielding samples at three different oleoresin-producing periods in this study. The WGCNA analysis was carried out to identify hub genes and gene modules that are highly correlated to oleoresin yields. As indicated in Figure 3A, 23 modules were subgrouped, of which genes had similar co-expression patterns. The number of hub genes within each candidate module ranged from 35 (steel blue) to 550 (blue) (Supplementary Materials Table S7). The hierarchical clustering of 1000 randomly selected genes by using the topological overlap matrix (TOM) indicates a relative high level of independence among these modules (Figure S2). Furthermore, the module–trait associations were also analyzed. The relationships of 23 functional modules and resin constitutive terpenoids were depicted in heat maps with corresponding correlation and p values (Figure 3B and Table S8). Focusing on the total oleoresin yield, the MEblue module exhibited the highest positive correlation ($r = 0.71$; $p < 0.001$), while the MEgreen module exhibited the highest negative correlation ($r = -0.67$; $p < 0.001$) (Figure 3B). As shown in Figure S3, the percentage of genes of the MEblue module and all genes were categorized into three GO pathways with the items of "biological process", "cellular component", and "molecular function".

Figure 3. WGCNA analysis of the transcripts associated with the biosynthetic and metabolic processes of pine oleoresin. (**A**) The hierarchical cluster tree of 30,330 unigenes. (**B**) The analysis of module–trait correlations, with rows representing the modules while columns representing the specific chemical compounds. The intersections of rows and columns are color-coded to represent the correlations based on the color legend.

3.6. Screening of Genes Related to Monoterpene Synthesis in High and Low Oleoresin-Yielding Slash Pine

The α- and β-pinene are two representatives of monoterpenes in pine resin and account for more than 40% of total oleoresin produced by high- and low-yielding slash pines. As indicated in Figure 3B, the MEtan, MEdarkgreen, and MEblue modules were the highest clusters that were positively correlated with α-pinene content, while the MEsteelblue and MElightgreen modules exhibited the highest positive correlation with the β-pinene content. Moreover, the eigengene adjacency heatmap indicated that there are high positive correlations among α-pinene content-related modules (MEtan, MEdarkgreen, and MEblue modules) and between MEsteelblue and MElightgreen modules in relation to β-pinene content (Figure 4A). Furthermore, the GO classification results of DEG exhibited that the genes of α- or β-pinene content-related modules were enriched in the "lipid transport and metabolism" and "transcription factor activity" pathways (Figure 4B). These outcomes suggest that transcription factors are of great importance in regulating the synthesis of α-pinene and β-pinene and their proportions in total oleoresin.
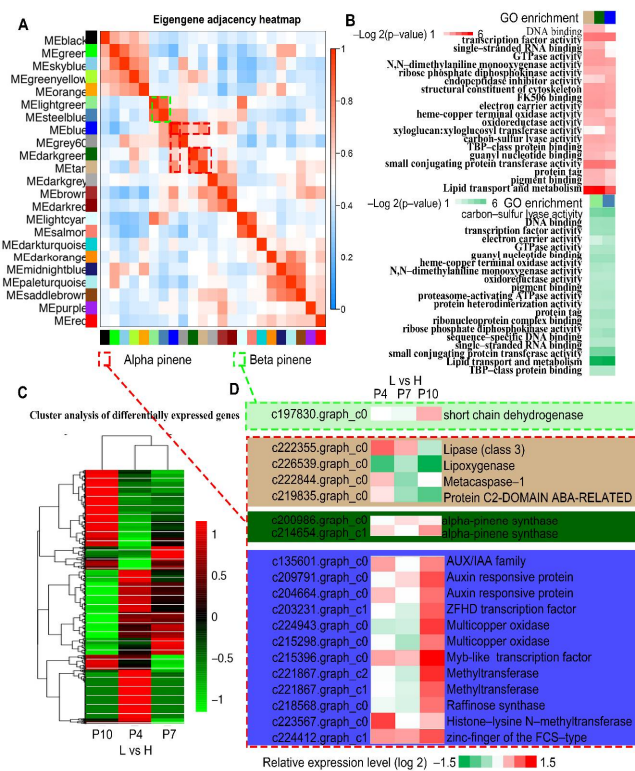

Figure 4. The DEGs associated with the biosynthetic and metabolic processes of α- and β-pinenes. (**A**) The eigengene adjacency heatmap of 23 modules. The dotted red or green frames represent co-expression modules associated with α- or β-pinenes contents, respectively. (**B**) GO enrichment analysis of genes in MEblue, MEdarkgreen, and MEtan modules associated with α-pinene, and MElightgreen, MEsteelblue associated with β-pinene. (**C**) Relative expression profiles of all genes in MEblue, MEdarkgreen, MEtan, MElightgreen, MEsteelblue modules. (**D**) Expression profiles of DEGs genes in MEblue, MEdarkgreen, MEtan, MElightgreen modules based on the RNA-seq data.

The relative genetic displays of the five modules are presented in Figure 4C using congregated analysis. Genes exhibiting different relative expressions at P4, P7, and P10 between high and low oleoresin-yielding slash pines were identified in these five modules, including one short-chain dehydrogenase gene (*c197830.graph_c0*) in the "MElightgreen" module; four genes (*c222355.graph_c0*, *c226539.graph_c0*, *c222844.graph_c0*, *c219835.graph_c0*) in the "MEtan" module; two α-pinene synthase genes (*c200986.graph_c0*, *c214654.graph_c1*) in the "MEdarkgreen" module; and 12 genes (*c135601.graph_c0*, *c209791.graph_c0*, *c204664.graph_c0*,

c203231.graph_c1, *c224943.graph_c0*, *c215298.graph_c0*, *c215396.graph_c0*, *c2221867.graph_c2*, *c221867.graph_c1*, *c218568.graph_c0*, *c223567.graph_c0*, *c224412.graph_c1*) in the "MEblue" module (Figure 4D).

3.7. Transcription Factors Involved in the Regulation of Pine Oleoresin Biosynthesis

In this study, functional annotation of unigenes identified a large number of transcription factors that are catagorized to diverse gene families, such as C2H2, basic helix–loop–helix (bHLH), *AP2*, MYB-related, and basic leucine zipper (bZIP) gene families, among others (Figure 5A). There were 20 transcription factors (TFs) annotated in the α-pinene content related modules (MEblue, MEdarkgreen, and MEtan), including four bHLH TFs, eight *AP2* TFs, five MYB TFs, one bZIP TFs and two WRKY TFs (Figure 5A,B). The relative expression levels of the 20 TFs in low and high oleoresin-yielding slash pines at P4, P7, and P10 are exhibited in Figure 5B. Furthermore, seven TFs with significant relative expression differences were validated by qRT-PCR assays (Figure 5B,C). The results of qRT-PCR indicated that the bHLH TF (*c211309.graph_c0*) and the WRKY TF (*c225743.graph_c0*) had lower expression levels in low-yielding slash pine than in high-yielding slash pine, while the other five TFs, including two *AP2* TFs (*c183227.graph_c0*; *c202525.graph_c0*), one bHLH TF (*c205514.graph_c0*), and two MYB TFs (*c212980.graph_c0*; *c215396.graph_c0*), had higher expression levels in low-yielding slash pine than in high-yielding ones (Figure 5C). These results were coincident with the findings that high-yielding slash pine has lower α-pinene content than low-yielding slash pine at P4, P7, and P10 (Figure 2C). In summary, TFs play a key role in regulating oleoresin biosynthesis, especially α-pinene biosynthesis in slash pine. Additionally, the differences in α-pinene content between high- and low-yielding slash pines are attributed to the variation in transcriptional levels of TFs.
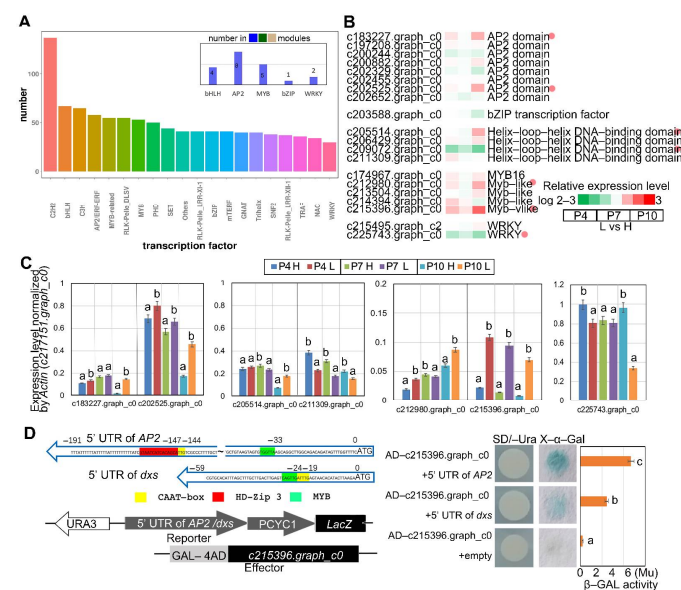
Figure 5. Transcript factors associated with α- and β-pinene biosynthesis. (**A**) Histogram of transcription factor family members annotated based on the unreferenced transcriptome data and 3 co-expression modules (MEblue, MEdarkgreen, MEtan) related to pinene content. (**B**) Expression profiles of *AP2*, bZIP, MYB and WRKY transcription factors as per the RNA-seq data. (**C**) Results of RT-PCR about the selected *AP2*, bZIP, MYB and WRKY transcription factors. The expression levels were standardized by the expression level of Actin (*c217151.graph_c0*). Four biological duplicates were showed in the experiment. (**D**) The 5′ UTR fragments of *AP2* (*c183227.graph_c0*) and *dxs* (*c193143.graph_c1*) were constructed in the pLacZi plasmid, which were co-transformed with GAL4-AD-MYB (*c215396.graph_c0*) plasmid into YM4271 strain. The implementation of X-gal staining with three technical replicates provided proof of the β-galactosidase activity. All bars represent means ± s.d. The columns labeled with different letters indicate a significant difference in *p* value < 0.05 according to Duncan's test.

3.8. Identification of the Candidate Genes Targeted by c215396.Graph_c0

According to the WGCNA analysis and relative expression analysis of TFs, the MYB TF, *c215396.graph_c0* is the most significant DEG between low- and high-yielding slash pines (Figure 5B,C). This finding suggested that *c215396.graph_c0* is one of the hub genes in regulating oleoresin biosynthesis. To further explore the downstream target genes of *c215396.graph_c0*, 5′ UTR sequences were extracted based on the RNA-seq data. The upstream sequences of *AP2* (*c183227.graph_c0*, −191 bp to ATG) and *dxs* (*c193143.graph_c1*, −59 bp to ATG) were successfully assembled. According to cis-acting element (CRE) analysis, the 5′ UTR sequences have the corresponding elements that can be recognized by MYB TFs (Figure 5D). Finally, the yeast one-hybrid (Y1H) assay was performed to explore the transcriptional activation activity of *c215396.graph_c0* on *AP2* and *dxs* (Figure 5D). The DNA sequence of *c215396.graph_c0* was constructed into a pGADT7-AD vector, and the 5′ UTR sequences of *AP2* and *dxs* were constructed into pLacZi vectors. The results of the yeast-one hybrid assay indicated that the MYB TF *c215396.graph_c0* can activate the transcription of *AP2* and *dxs*.

4. Discussion

Oleoresin yield is a highly heritable trait that is genetically controlled by quantitative trait loci (QTLs) [36,37]. To date, little is known about the underlying genetic mechanism of oleoresin biosynthesis in slash pine. In the present study, 20 terpenoids were identified in the oleoresin of slash pine using GC-MS, and the major terpenoid compositions are exclusively composed of monoterpenes and diterpenes. Among monoterpenes, α- and β-pinene account for the majority, while myrcene, dipentene, and camphene make up less than 0.1% of the total monoterpenes. The diterpenes are mainly composed of isopimaric acid, abietic acid, levopimaric acid, dehydroabietic acid, and neoabietic acid. Intriguingly, there are no sesquiterpenes (more than 0.1%) identified in oleoresin of slash pine, whereas sesquiterpenes are one kind of major constituent of turpentine oil in oleoresin from other pine species, such as Masson pine and Aleppo pine [38,39]. The proportions of α-pinene, β-pinene, abietic acid, levopimaric acid, and dehydroabietic acid of oleoresin from slash pine exhibited significant variation at P4, P7, and P10. The proportion of α-pinene reached the highest level in oleoresin produced in July, while β-pinene, abietic acid, levopimaric acid, and dehydroabietic acid occupied greater percentages in oleoresin harvested in October than in any other months. Moreover, the proportion of these five terpenoids in oleoresin of high- and low-yielding slash pines showed remarkable differences.

The KEGG enrichment results showed that a total of 34 DEGs of L vs. H were enriched in pathways related to terpenoid backbone biosynthesis, including diterpenoid biosynthesis, carotenoid biosynthesis, and ubiquinone and other terpenoid biosynthesis pathways. In the MEP pathway, the chloroplast-located *dxs* and *ispF/G* are rate-limiting enzymes play pivotal roles in synthesizing 1-hydroxy-2-methyl-2-butenyl 4-diphosphate [40–43]. The homologues of *dxs* and *ispF/G* in slash pine were significantly highly induced in the high-yielding trees compared with the low-yielding trees. However, enzymes in the mevalonate pathway, including the cytoplasm-located mevalonate kinase, peroxisome-located phosphomevalonate kinase, and diphosphomevalonate decarboxylase [44–46], showed nuanced differences between high- and low-yielding slash pines. The intermediate monoprenyl diphosphate is a crucial substrate for the biosynthesis of isoprene under the catalyzation of *ispS* [47]. The upregulated expression of *ispS* in low-yielding pine was consistent with the fact that oleoresin of low-yielding pine had more isoprene. As for TPS, the homologous proteins of *PT30* and *QH6* in slash pine catalyzed the synthesis of α- and β-pinene, respectively. Moreover, homologous proteins of *SPS*, *ZFPS*, and *idsA* are responsible for the synthesis of ubiquinone, sesquiterpenoid, and diterpenoid, respectively. According to the previous studies, the homology of *PT30* in Loblolly pine produces both (+)-α-pinene and (−)-α-pinene in a 97:3 ratio [48], and *QH6* can convert GPP to (−)-β-pinene and (−)-α-pinene in a 94:6 ratio in Abies grandis and Artemisia annua [49,50]. Therefore, homologies of *PT30* and *QH6* in slash pine are also implicated in controlling not only the

content of but also the ratio between α- and β-pinenes. The expression levels of *PT30* and *QH6* in high- and low-yielding pines coincided with the content of α- and β-pinenes in the slash pine. Orthologs of plastid-located *SPS* in Arabidopsis are involved in the biosynthesis of plastoquinone-9 (PQ-9) [51] and can improve the tolerance of these plants to excess light energy [52]. Multiple sesquiterpenes in Solanum lycopersicum and Solanum habrochaites have been found to be associated with *ZFPS* [53,54]. Although the homology of *ZFPS* had differential expression level between high- and low-yielding slash pines, terpenoids of sesquiterpenes were hardly detected in oleoresin of slash pine according to the data from GC-MS. GGPP, as an intermediate product of the terpenoid backbone biosynthesis pathway, is precursor of gibberellin (GAs) [55] and diterpenes such as the side chain of chlorophylls [56]. The lower expression of *idsA* in high-yielding pines suggested that the content of GAs and chlorophyll in these plants may be lower than that of low-yielding pines, which is consistent with the results in *Pinus massoniana* [57]. Intriguingly, low-yielding slash pine had a lower growth rate than the high-yielding pines, as the average heights of the high- and low-yielding slash pines were 14.32 ± 1.12 m and 11.71 ± 1.23 m, respectively.

Previous studies in Arabidopsis thaliana have provided a comprehensive and systematic overview of various terpenoid biosynthesis pathways. In order to identify the core gene modules and hub genes in α- and β-pinene biosynthesis, WGCNA analysis was performed. After GO enrichment analysis, modules of transcription factor activity and lipid transport and metabolism exhibited closed correlations with biosynthesis of monoterpenes. Therefore, TFs were verified to be crucial for the biosynthetic process of α- and β-pinene. According to the annotation, 20 TFs of 5 TF families were identified in "MEdarkgreen" and the related "MEdarkgreen" and "MEtan" modules that are associated with α- and β-pinene biosynthesis. The MYB TF gene (*c215396.graph_c0*). is most differentially expressed between high- and low-yielding slash pines. We postulated that the downstream targets should be DEGs of L vs. H. According to the homologous gene sequences in Arabidopsis and the potential cis-acting element analysis based on the online PlantCARE webserver [58], only two DEGs (*AP2* and *dxs*) were identified with their 5′UTR regions harboring MYB TF recognition sites and CAAT-boxes (RNA polymerase binding sites) [59]. The subsequent one-hybrid yeast assay verified that *c215396.graph_c0* can recognise the ATG upstream sequences of *AP2* and *dxs*. The homologous gene of *AP2* (*c183227.graph_c0*) in Catharanthus roseus is ORCA3, which has been reported to regulate the biosynthetic process of terpenoid indole alkaloids (TIAs) [60–62]. In general, the MYB TF (*c215396.graph_c0*) directly regulates transcription of *dxs* in the MEP pathway and indirectly affects TIA synthesis via the *AP2* (Figure 6).
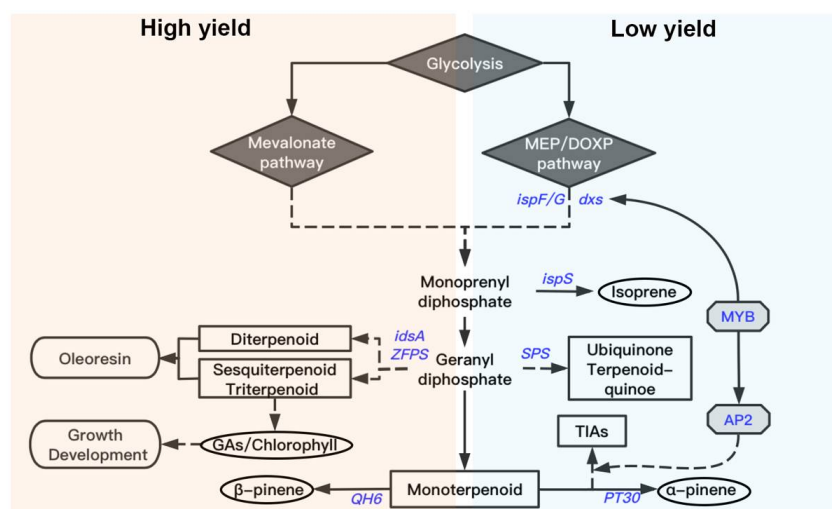
Figure 6. The proposed model of oleoresin biosynthesis in slash pine.

5. Conclusions

In the present study, the terpenoid compositions of oleoresin in high and low resin-yielding slash pines were explored, and monoterpenes and diterpenes were identified to be the major components. The α- and β-pinenes make up the majority of monoterpenes and abietic acid, levopimaric acid, and neoabietic acid are most abundant in diterpenes. However, these terpenoids occupy different proportions in oleoresin produced by high- and low-yielding slash pine. Moreover, the RNA-seq analysis, depending on the secondary xylems of high- and low-yielding pines, provides a comprehensive transcriptional profile for future research. In RNA-seq data, a series of DEGs associated with terpenoid backbone biosynthesis were characterized, including *ispF/G/S*, *dxs*, *SPS*, *PT30*, *idsA*, *ZFPS*, *QH6*, etc., which are widely implicated in the MEP pathway and downstream terpenoid biosynthesis processes. Furthermore, the WGCNA analysis and subsequent one-hybrid yeast assay identified a MYB TF gene (*c215396.graph_c0*). *c215396.graph_c0* is a key gene that functions as a transcriptional activator regulating the expression of *dxs* and *AP2* (Figure 6). These genes contribute to the differential α-pinene content in high- and low-yielding pines. In summary, our work identified and characterized the crucial genes and gene modules involving in terpenoid biosynthesis that constitute a complicated gene regulation network differentially controlling the oleoresin yield and composition in high- and low-yielding slash pines.

References

1. Phillips, M.A.; Croteau, R.B. Resin-based defenses in conifers. *Trends Plant Sci.* **1999**, *4*, 184–190. [CrossRef]
2. Pichersky, E.; Gershenzon, J. The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. *Curr. Opin. Plant Biol.* **2002**, *5*, 237–243. [CrossRef]
3. Neis, F.A.; de Costa, F.; de Almeida, M.R.; Colling, L.C.; Junkes, C.F.D.O.; Fett, J.P.; Fett-Neto, A.G. Resin exudation profile, chemical composition, and secretory canal characterization in contrasting yield phenotypes of Pinus elliottii Engelm. *Ind. Crop. Prod.* **2019**, *132*, 76–83. [CrossRef]

4. Loreto, F.; Pinelli, P.; Manes, F.; Kollist, H. Impact of ozone on monoterpene emissions and evidence for an isoprene-like antioxidant action of monoterpenes emitted by Quercus ilex leaves. *Tree Physiol.* **2004**, *24*, 361–367. [CrossRef]
5. Sharkey, T.D.; Yeh, S. ISOPRENE EMISSION FROM PLANTS. *Annu. Rev. Plant Biol.* **2001**, *52*, 407–436. [CrossRef]
6. Ramawat, K.G.; Mérillon, J.M. *Natural products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes*; Springer: Berlin/Heidelberg, Germany, 2013. [CrossRef]
7. Sharon, H. Book Reviews. *Act. Learn. High. Educ.* **2001**, *2*, 81–82. [CrossRef]
8. Demyttenaere, J.; De Kimpe, N. Biotransformation of terpenes by fungi: Study of the pathways involved. *J. Mol. Catal. B Enzym.* **2001**, *11*, 265–270. [CrossRef]
9. Chatzivasileiou, A.O.; Ward, V.; Edgar, S.M.; Stephanopoulos, G. Two-step pathway for isoprenoid synthesis. *Proc. Natl. Acad. Sci. USA* **2018**, *116*, 506–511. [CrossRef]
10. Tholl, D. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* **2006**, *9*, 297–304. [CrossRef]
11. Liang, P.-H. Reaction kinetics, catalytic mechanisms, conformational changes, and inhibitor design for prenyltransferases. *Biochemistry* **2009**, *48*, 6562–6570. [CrossRef]
12. Liang, P.-H.; Ko, T.-P.; Wang, A.H.-J. Structure, mechanism and function of prenyltransferases. *JBIC J. Biol. Inorg. Chem.* **2002**, *269*, 3339–3354. [CrossRef]
13. Gershenzon, J.; Kreis, W. Biochemistry of Terpenoids: Monoterpenes, Sesquiterpenes, Diterpenes, Sterols, Cardiac Glycosides and Steroid Saponins. *Biochem. Plant Second. Metab.* **2018**, *2*, 218–294. [CrossRef]
14. Schmidt, A.; Waⅇchter, B.; Temp, U.; Krekling, T.; Seⅇguin, A.; Gershenzon, J. A Bifunctional Geranyl and Geranylgeranyl Diphosphate Synthase Is Involved in Terpene Oleoresin Formation in *Picea abies*. *Plant Physiol.* **2009**, *152*, 639–655. [CrossRef]
15. Trapp, S.; Croteau, R. DEFENSIVE RESIN BIOSYNTHESIS IN CONIFERS. *Annu. Rev. Plant Biol.* **2001**, *52*, 689–724. [CrossRef]
16. Trapp, S.C.; Croteau, R.B. Genomic Organization of Plant Terpene Synthases and Molecular Evolutionary Implications. *Genetics* **2001**, *158*, 811–832. [CrossRef]
17. Yi, W.; Utilization, Y.A.O.F.K.L.O.F.P.C.A.; Xiaolong, Y.; Mei, H.; Jiang, L.; Juan, W. Transcriptome and gene expression analysis revealed mechanisms for producing high oleoresin yields from Simao pine (*Pinus kesiya* var. *langbianensis)*. *Plant Omics* **2018**, *11*, 42–49. [CrossRef]
18. Liu, Q.; Zhou, Z.; Wei, Y.; Shen, D.; Feng, Z.; Hong, S. Genome-Wide Identification of Differentially Expressed Genes Associated with the High Yielding of Oleoresin in Secondary Xylem of Masson Pine (Pinus massoniana Lamb) by Transcriptomic Analysis. *PLoS ONE* **2015**, *10*, e0132624. [CrossRef]
19. Mei, L.; Li, Z.; Yan, Y.; Wen, Z.; Wen, X.; Yang, Z.; Feng, Y. Identification and functional study of oleoresin terpenoid biosynthesis-related genes in masson pine (*Pinus massoniana* L.) based on transcriptome analysis. *Tree Genet. Genomes* **2020**, *16*, 53. [CrossRef]
20. Rodrigues, K.; Azevedo, P.; Sobreiro, L.; Pelissari, P.; Fett-Neto, A. Oleoresin yield of Pinus elliottii plantations in a subtropical climate: Effect of tree diameter, wound shape and concentration of active adjuvants in resin stimulating paste. *Ind. Crop. Prod.* **2008**, *27*, 322–327. [CrossRef]
21. Susaeta, A.; Peter, G.F.; Hodges, A.W.; Carter, D.R. Oleoresin tapping of planted slash pine (*Pinus elliottii* Engelm. var. *elliottii*) adds value and management flexibility to landowners in the southern United States. *Biomass-Bioenergy* **2014**, *68*, 55–61. [CrossRef]
22. Falkenhagen, E.R. Relationships between some genetic parameters and test environments in open-pollinated families of *Pinus elliottii* in South Africa. *Theor. Appl. Genet.* **1989**, *77*, 857–866. [CrossRef]
23. Barnett, J.P.; Sheffield, R.M. Slash Pine: Characteristics, History, Status and Trends. Slash Pine: Still Growing and Growing! In Proceedings of the Slash Pine Symposium, Asheville, NC, USA, April 23–25 2002.
24. Lange, B.M. The Evolution of Plant Secretory Structures and Emergence of Terpenoid Chemical Diversity. *Annu. Rev. Plant Biol.* **2015**, *66*, 139–159. [CrossRef]
25. Yi, M.; Jia, T.; Dong, L.; Zhang, L.; Leng, C.; Liu, S.; Lai, M. Resin yield in *Pinus elliottii* Engelm is related to the resin flow rate, resin components and resin duct characteristics at three locations in southern China. *Ind. Crop. Prod.* **2020**, *160*, 113141. [CrossRef]
26. Cunningham, A. Pine resin tapping techniques used around the world. *Pine Resin Biol. Chem. Appl.* **2012**, *661*, 1–8.
27. Hodges, A.W.; Johnson, J.D. Borehole Oleoresin Production from Slash Pine. *South. J. Appl. For.* **1997**, *21*, 108–115. [CrossRef]
28. Junkes, C.F.D.O.; Duz, J.V.V.; Kerber, M.R.; Wieczorek, J.; Galvan, J.L.; Fett, J.P.; Fett-Neto, A.G. Resinosis of young slash pine (*Pinus elliottii* Engelm.) as a tool for resin stimulant paste development and high yield individual selection. *Ind. Crop. Prod.* **2019**, *135*, 179–187. [CrossRef]
29. Junkes, C.F.D.O.; Júnior, A.T.D.A.; de Lima, J.C.; de Costa, F.; Füller, T.; de Almeida, M.R.; Neis, F.A.; Rodrigues-Corrêa, K.C.D.S.; Fett, J.P.; Fett-Neto, A. Resin tapping transcriptome in adult slash pine (*Pinus elliottii* var. *elliottii*). *Ind. Crop. Prod.* **2019**, *139*, 111545. [CrossRef]
30. Lai, M.; Zhang, L.; Lei, L.; Liu, S.; Jia, T.; Yi, M. Inheritance of resin yield and main resin components in *Pinus elliottii* Engelm. at three locations in southern China. *Ind. Crop. Prod.* **2019**, *144*, 112065. [CrossRef]
31. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **2013**, *29*, 644–652. [CrossRef]
32. Mao, X.; Cai, T.; Olyarchuk, J.G.; Wei, L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **2005**, *21*, 3787–3793. [CrossRef]

33. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef]

34. Allen, G.C.; Flores-Vergara, M.A.; Krasynanski, S.; Kumar, S.; Thompson, W. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **2006**, *1*, 2320–2325. [CrossRef]

35. Liu, C.; Wang, B.; Li, Z.; Peng, Z.; Zhang, J. TsNAC1 Is a Key Transcription Factor in Abiotic Stress Resistance and Growth. *Plant Physiol.* **2017**, *176*, 742–756. [CrossRef]

36. Zhang, S.; Jiang, J.; Luan, Q. Genetic and correlation analysis of oleoresin chemical components in slash pine. *Genet. Mol. Res.* **2016**, *15*, 1–12. [CrossRef]

37. Roberds, J.H.; Strom, B.L.; Hain, F.P.; Gwaze, D.P.; McKeand, S.; Lott, L.H. Estimates of genetic parameters for oleoresin and growth traits in juvenile loblolly pine. *Can. J. For. Res.* **2003**, *33*, 2469–2476. [CrossRef]

38. Karanikas, C.; Walker, V.; Scaltsoyiannes, A.; Comte, G.; Bertrand, C. High vs. low yielding oleoresin Pinus halepensis Mill. trees GC terpenoids profiling as diagnostic tool. *Ann. For. Sci.* **2010**, *67*, 412. [CrossRef]

39. Bai, Q.; Zhang, Q.; Cai, Y.; Lian, H.; Liu, W.; Luo, M.; Zeng, L.; He, B. Genome-wide association study of terpenoids in resin reveals candidate genes for resin yield in Pinus massoniana. *Dendrobiology* **2020**, *84*, 109–121. [CrossRef]

40. Estévez, J.M.; Cantero, A.; Reindl, A.; Reichler, S.; León, P. 1-Deoxy-d-xylulose-5-phosphate Synthase, a Limiting Enzyme for Plastidic Isoprenoid Biosynthesis in Plants. *J. Biol. Chem.* **2001**, *276*, 22901–22909. [CrossRef]

41. Hsieh, M.-H.; Goodman, H.M. Functional evidence for the involvement of Arabidopsis *IspF* homolog in the nonmevalonate pathway of plastid isoprenoid biosynthesis. *Planta* **2005**, *223*, 779–784. [CrossRef]

42. Hsieh, M.-H.; Chang, C.-Y.; Hsu, S.-J.; Chen, J.-J. Chloroplast localization of methylerythritol 4-phosphate pathway enzymes and regulation of mitochondrial genes in ispD and ispE albino mutants in Arabidopsis. *Plant Mol. Biol.* **2008**, *66*, 663–673. [CrossRef]

43. Querol, J.; Campos, N.; Imperial, S.; Boronat, A.; Rodríguez-Concepción, M. Functional analysis of the *Arabidopsis thaliana* GCPE protein involved in plastid isoprenoid biosynthesis. *FEBS Lett.* **2002**, *514*, 343–346. [CrossRef]

44. Simkin, A.J.; Guirimand, G.; Papon, N.; Courdavault, V.; Thabet, I.; Ginis, O.; Bouzid, S.; Giglioli-Guivarc'H, N.; Clastre, M. Peroxisomal localisation of the final steps of the mevalonic acid pathway in planta. *Planta* **2011**, *234*, 903–914. [CrossRef]

45. Cordier, H.; Karst, F.; Bergès, T. Heterologous expression in Saccharomyces cerevisiae of an Arabidopsis thaliana cDNA encoding mevalonate diphosphate decarboxylase. *Plant Mol. Biol.* **1999**, *39*, 953–967. [CrossRef]

46. Henry, L.K.; Gutensohn, M.; Thomas, S.T.; Noel, J.P.; Dudareva, N. Orthologs of the archaeal isopentenyl phosphate kinase regulate terpenoid production in plants. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10050–10055. [CrossRef]

47. Sasaki, K.; Ohara, K.; Yazaki, K. Gene expression and characterization of isoprene synthase from *Populus alba*. *FEBS Lett.* **2005**, *579*, 2514–2518. [CrossRef]

48. Phillips, M.A.; Wildung, M.R.; Williams, D.C.; Hyatt, D.C.; Croteau, R. cDNA isolation, functional expression, and characterization of (+)-α-pinene synthase and (−)-α-pinene synthase from loblolly pine (Pinus taeda): Stereocontrol in pinene biosynthesis. *Arch. Biochem. Biophys.* **2003**, *411*, 267–276. [CrossRef]

49. Lu, S.; Xu, R.; Jia, J.-W.; Pang, J.; Matsuda, S.P.; Chen, X.-Y. Cloning and Functional Characterization of a β-Pinene Synthase from *Artemisia annua* That Shows a Circadian Pattern of Expression. *Plant Physiol.* **2002**, *130*, 477–486. [CrossRef]

50. Bohlmann, J.; Steele, C.L.; Croteau, R. Monoterpene Synthases from Grand Fir (Abies grandis). *J. Biol. Chem.* **1997**, *272*, 21784–21792. [CrossRef]

51. Block, A.; Fristedt, R.; Rogers, S.; Kumar, J.; Barnes, B.; Barnes, J.; Elowsky, C.G.; Wamboldt, Y.; Mackenzie, S.A.; Redding, K.; et al. Functional Modeling Identifies Paralogous Solanesyl-diphosphate Synthases That Assemble the Side Chain of Plastoquinone-9 in Plastids. *J. Biol. Chem.* **2013**, *288*, 27594–27606. [CrossRef]

52. Ksas, B.; Becuwe, N.; Chevalier, A.; Havaux, M. Plant tolerance to excess light energy and photooxidative damage relies on plastoquinone biosynthesis. *Sci. Rep.* **2015**, *5*, 10919. [CrossRef]

53. Sallaud, C.; Rontein, D.; Onillon, S.; Jabès, F.; Duffé, P.; Giacalone, C.; Thoraval, S.; Escoffier, C.; Herbette, G.; Leonhardt, N.; et al. A Novel Pathway for Sesquiterpene Biosynthesis from *Z,Z*-Farnesyl Pyrophosphate in the Wild Tomato *Solanum habrochaites*. *Plant Cell* **2009**, *21*, 301–317. [CrossRef]

54. Akhtar, T.A.; Matsuba, Y.; Schauvinhold, I.; Yu, G.; Lees, H.A.; Klein, S.E.; Pichersky, E. The tomato *cis*-prenyltransferase gene family. *Plant J.* **2012**, *73*, 640–652. [CrossRef]

55. Brock, N.L.; Tudzynski, B.; Dickschat, J.S. Biosynthesis of Sesqui- and Diterpenes by the Gibberellin Producer Fusarium fujikuroi. *ChemBioChem* **2011**, *12*, 2667–2676. [CrossRef]

56. Ruiz-Sola, M.; Coman, D.; Beck, G.; Barja, M.V.; Colinas, M.; Graf, A.; Welsch, R.; Rütimann, P.; Bühlmann, P.; Bigler, L.; et al. *Arabidopsis* GERANYLGERANYL DIPHOSPHATE SYNTHASE11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. *New Phytol.* **2015**, *209*, 252–264. [CrossRef]

57. Yijiang, H.; Bangping, C.; Yichi, L. A Study on the Relationships between Chlorophyll Content and Resin Productivity of Masson Pine. *J. Fujian Coll. For.* **1998**, *18*, 58. [CrossRef]

58. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [CrossRef]

59. Farrell, R.E. Transcription and the Organization of Eukaryotic Genes. In *RNA Methodologies*; Elsevier Academic Press: Cambridge, MA, USA, 2005.

60. De Boer, K.; Tilleman, S.; Pauwels, L.; Bossche, R.V.; De Sutter, V.; Vanderhaeghen, R.; Hilson, P.; Hamill, J.D.; Goossens, A. apetala2/ethylene response factor and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. *Plant J.* **2011**, *66*, 1053–1065. [CrossRef]

61. Montiel, G.; Zarei, A.; Körbes, A.P.; Memelink, J. The Jasmonate-Responsive Element from the ORCA3 Promoter from Catharanthus roseus is Active in Arabidopsis and is Controlled by the Transcription Factor AtMYC2. *Plant Cell Physiol.* **2011**, *52*, 578–587. [CrossRef]

62. Wang, C.-T.; Liu, H.; Gao, X.-S.; Zhang, H.-X. Overexpression of G10H and ORCA3 in the hairy roots of Catharanthus roseus improves catharanthine production. *Plant Cell Rep.* **2010**, *29*, 887–894. [CrossRef]