Genomic-Wide Identification and Characterization of the Uridine Diphosphate Glycosyltransferase Family in *Eucommia ulmoides* Oliver

Dan Ouyang [1,2], Lan-Chun Wang [1], Ting Tang [1] and Hong Feng [1,2,*]

1   Key Laboratory of Molecular Biology and Biotechnology of Sichuan Province, Department of Biotechnology, College of Life Sciences, Sichuan University, Chengdu 610064, China; ouyangd@stu.scu.edu.cn (D.O.); wanglc@stu.scu.edu.cn (L.-C.W.); tangting@stu.scu.edu.cn (T.T.)
2   Key Laboratory for Bio-Resources and Eco-Environment of Ministry of Education, Department of Biotechnology, Sichuan University, 29 Wangjiang Road, Chengdu 610064, China
*   Correspondence: hfeng@scu.edu.cn

Abstract: *Eucommia ulmoides* Oliver is a woody plant with great economic and medicinal value. Its dried bark has a long history of use as a traditional medicinal material in East Asia, which led to many glycosides, such as aucubin, geniposide, hyperoside, astragalin, and pinoresinol diglucoside, being recognized as pharmacologically active ingredients. Uridine diphosphate glycosyltransferases (UGTs) catalyze a glycosyl-transferring reaction from the donor molecule uridine-5′-diphosphate-glucose (UDPG) to the substrate, which plays an important role in many biological processes, such as plant growth and development, secondary metabolism, and environmental adaptation. In order to explore the biosynthetic pathways of glycosides in *E. ulmoides*, 91 putative *Eu*UGT genes were identified throughout the complete genome of *E. ulmoides* through function annotation and an UDPGT domain search. Phylogenetic analysis categorized them into 14 groups. We also performed GO annotations on all the *Eu*UGTs to gain insights into their functions in *E. ulmoides*. In addition, transcriptomic analysis indicated that most *Eu*UGTs showed different expression patterns across diverse organs and various growing seasons. By protein–protein interaction predication, a biosynthetic routine of flavonoids and their glycosides was also proposed. Undoubtedly, these results will help in future research into the biosynthetic pathways of glycoside compounds in *E. ulmoides*.

Keywords: *Eucommia ulmoides*; UDP-glycosyltransferase; transcriptome; expression; lignan; flavonoid; biosynthesis

1. Introduction

Glycosyltransferases (GT, EC 2.4.x.y) catalyze to transfer a glycosyl moiety from a donor molecule to an acceptor substrate to form a glycosidic bond, which is employed by cells for the biosynthesis of glycolipids, glycoproteins, hormones, and various glycosides. GTs are ubiquitous in all kinds of organisms, from bacteria to animals. To date, 114 families of glycosyltransferases containing more than 850,000 members are recognized and classified in the CAZy database (http://www.cazy.org, accessed on 5 August 2021) [1,2], among which, the largest family (GT1) transfers sugars onto numerous small molecules.

In plants, GTs typically utilize uridine diphosphate (UDP)-activated sugars as donor molecules, by which glycosylation reactions not only synthesize essential substances for growth, but also glycosylate various secondary metabolites, such as monolignols, anthocyanins, and terpenoids [3,4]. Plant UDP-glycosyltransferases (UGTs) have a conserved motif of 44 residues called PSPG box (putative secondary plant glycosyltransferase box) in the *C*-terminal, which is featured as a UDP-sugar binding domain [5]. In light of their important roles in many cellular processes, genomic-wide identification and phylogenetic analyses of UGTs were carried out for many plant species. For example, 123, 147, 179, 180 and 137 UGTs have been identified in *Arabidopsis thaliana* [6], *Zea mays* [7], *Triticum*

aestivum [8], *Oryza sativa* [9], and *Linum usitatissimum* [10], respectively. Through phylogenetic analyses, plant UGTs were classified into 18 groups (A to R) [6,11], indicating the significant divergence and expansion of the UGT family during evolution [12]. Moreover, identification of the UGT gene family is also helpful for screening out the UGTs that may be included in specific metabolic pathways. For example, a combination of transcriptome and metabolite analyses of the adventitious roots treated with methyl-jasmonate led to the identification of several putative UGTs involved in ginsenoside biosynthesis in Korean ginseng (*Panax ginseng*) [13]. Through comparative transcriptome analysis between two closely related *Ilex* species, a UGT was recognized as a candidate for participation in triterpenoid biosynthesis [14]. In addition, sequence analysis of the UGT genes in *Camellia sinensis* allowed the identification of two genes encoding flavonol 3-*O*-glucosyltransferase and flavonol 3-*O*-galactosyltransferase [15].

Eucommia ulmoides Oliver is a kind of woody plant with high economic and remedial value, since its bark has a long history of use as a traditional medicine in East Asia [16]. In recent years, many studies were carried out on *E. ulmoides*, especially in pharmacology. It was found that the extracted components of *E. ulmoides* perform a wide range of pharmacological activities, such as serving as an anti-hypertension, anti-hyperlipidemia, anti-diabetic, anti-bacterial, anti-inflammation, and antioxidant agent, as well as protecting of the kidneys and liver from injury [17–19]. Various compounds with pharmacological potential from various tissues of *E. ulmoides* were subsequently identified, such as chlorogenic acid, geniposide, aucubin, rutin, hyperoside, astragalin, and pinoresinol diglucoside (PDG), etc. [16]. Many of these components are glycosides. For example, geniposide, a kind of bioactive iridoid glycoside, creates anti-inflammatory effects by reducing the release of inflammatory cytokines [20] and can also alleviate obesity-related cardiac injury by activating AMPK-α and the Sirt1 pathway [21]. Another glycoside of aucubin in *E. ulmoides* was shown to have anti-inflammatory effects, which may inhibit pulmonary inflammation and fibrosis in a mouse model established by the intratracheal injection of bleomycin [22]. Moreover, aucubin was also shown to have liver-protecting properties and to suppress DNA replication of HBV in vitro [23]. In addition, the PDG found in *E. ulmoides* is a natural hypotensive compound, which could protect against oxidative low-density lipoprotein-induced dysfunction in vein endothelial cells [24], thus becoming a promising drug for the treatment of cardiovascular disease.

Although many natural glycosides with pharmacological potential have been isolated and identified in *E. ulmoides*, almost none of the *Eu*UGTs have been biochemically characterized up until now. For instance, a *Eu*UGT cDNA was cloned and another *Eu*UGT was recombinantly expressed in *Escherichia coli* [25,26]. However, their catalytic activity and substrate were not definitively confirmed. In order to identify the candidate UGTs involved in the biosynthesis of glycosides in *E. ulmoides*, a genome-wide comprehensive analysis of *Eu*UGT genes should be the first step. The genome-wide identification of a gene family has been successfully applied to numerous gene families in plants. For example, the AT-Hook Motif Nuclear Localized gene family and β-Ketoacyl CoA synthetase gene family were identified in *Glycine max* L. [27] and *Hordeum vulgare* L. [28], respectively, from which the biochemical characteristics, gene structure, evolutionary relationships, and expression patterns of these gene families' members could be deduced. Furthermore, the assignment of biological functions to each member of the gene family is attractive. Through genome-wide analysis and comparative transcriptomics, the PIN-formed and brassinazole-resistant gene families were identified in wheat, which were recognized as playing crucial roles in response to diverse biotic and abiotic stresses [29,30]. In addition, a genome-wide strategy was also used to characterize long non-coding RNAs and natural antisense transcripts [31,32].

In this work, 91 UGTs were identified through a comprehensive search of the draft genome of *E. ulmoides*, and then compared with the UGTs of *A. thaliana* and *L. usitatissimum* L. based on a phylogenetic analysis. Further, the expression pattern of *Eu*UGTs was also analyzed using third-party transcriptome data available in a public database. Hopefully,

these data can promote future research on UDP-glycosyltransferase in order to elucidate their catalytic mechanism and the metabolic pathways of the glycosides in *E. ulmoides*.

2. Results

2.1. Genome-Wide Identification of Putative UGTs in E. ulmoides

By blasting the *E. ulmoides* draft genome [33] with the UDPGT domain (PF00201.18), 91 putative *Eu*UGT genes were identified and verified manually (Supplementary file S1 and S2), 81 of which were previously annotated as UGTs [33]. However, the other genes annotated as UDP-glycosyltransferase/glucosyltransferase, glycosyltransferase, or UGT were excluded due to the absence of a PSPG box. All of the putative *Eu*UGTs contained a conserved PSPG motif at their *C*-terminal. The organization of the 91 *Eu*UGT genes was also examined. It was shown that the majority of *Eu*UGT genes (48, 52.7%) had no intron. Among the *Eu*UGT genes containing introns, 37 out of 43 (86.05%) had only one intron. Four *Eu*UGT genes contained two introns and one gene had three introns (Figure S1).

The amino acid similarity between individual pairs of these sequences varied from 21% to 100%. Among them, three pairs of *Eu*UGTs (GWHPAAAL014233 and GWH-PAAAL012139; GWHPAAAL022859 and GWHPAAAL018716; GWHPAAAL025228 and GWHPAAAL008406) had 100% amino acid sequence identity and the same gene structure (Table S1), despite being localized in different scaffolds. The other 29 *Eu*UGTs were found to be located in tandem in the same scaffold with a high similarity of up to 73% in the amino acid sequence and almost identical gene structure (Table S1).

The protein sequence lengths and molecular weights of the *Eu*UGTs ranged from 261 (GWHPAAAL023819) to 569 (GWHPAAAL006284) amino acid residues, and 29.1 (GWHPAAAL23819) to 64.3 (GWHPAAAL006284) kDa, respectively. The estimations of theoretical pI ranged from 4.82 (GWHPAAAL014996) to 8.89 (GWHPAAAL005638). Out of the 91 *Eu*UGTs, 17 were predicted as stable proteins, and 74 proteins were predicted as unstable proteins. The grand average hydropathicities of 91 *Eu*UGTs were calculated to suggest that most of the *Eu*UGTs (83) are hydrophilic proteins. The subcellular locations of the 91 *Eu*UGTs were also predicated, with 63 in the cytoplasm, 14 in the chloroplast, 11 on the plasma membrane, 1 in extracellular space, 1 in the lysosome, and 1 in the mitochondria. Detailed information is summarized in Table S2.

2.2. Phylogenetic Analysis of EuUGTs

Originally, UGTs of the model plant *A. thaliana* were classified into 14 phylogenetic groups (A~N) based on the homology of their amino acid sequences [6]. Subsequently, four new groups (O, P, Q and R) were found in *P. ginseng* [13], *O. sativa* [9], and *Malus domestica* [34]. Based on this, 91 *Eu*UGTs were later clustered with the selected 246 UGTs (for more detail, see the Materials and Methods section) from other plants to construct a phylogenetic tree (Figure 1A). All of the UGTs mentioned above were divided into 16 groups (A-N, O and R), which are identical to the 18 known groups, excluding groups P and Q. The UGTs of groups P and Q were discovered in the other plants, such as *Punica granatum* L. [11], *T. aestivum* [8], and *Camellia sinensis* [15]. The *Eu*UGTs were categorized into 14 groups, such as A to E, G to M, and O and R. In the set of *Eu*UGTs analyzed, there were no *Eu*UGTs from groups F, N, Q, or P.
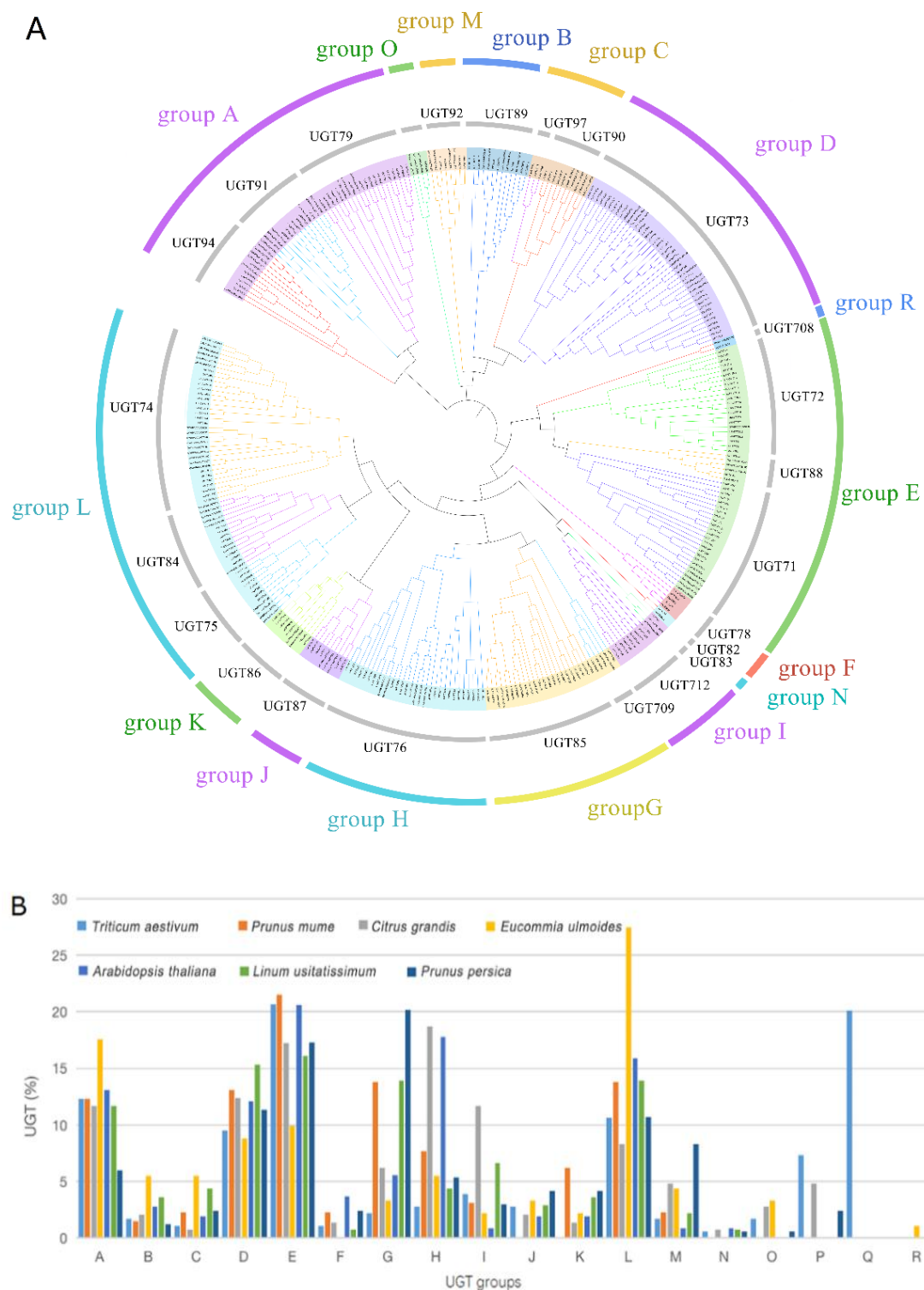
Figure 1. Phylogenetic tree of UGTs in E. ulmoides, L. usitatissimum and A. thaliana constructed by neighbor-joining method (**A**) and the distribution of members in the phylogenetic groups among selected plants species (**B**). In (**A**), each color arc corresponds to a phylogenetic group and each gray arc to a class of plant UGTs according to the UGT nomenclature system [5].

The distribution of UGTs in each group were compared among the selected plant species [12] (Figure 1B). Most of the *Eu*UGTs were categorized into the groups A, D, E, and L. The number of *Eu*UGTs in these groups are believed to expand rapidly during the evolution of higher plants [9,12]. In general, the UGTs of group E and L have a wide

range of substrates. Group E has been identified in most plants [9], indicating the existence of Group E is conservative during plant evolution. Group E mainly contains flavonoids glycosyltransferases, abscisic acid glycosyltransferases, and monolignol glycosyltransferases. Group L appears only in vascular plants [9]. Substrates of members of Group L contain auxin, anthranilate, anthocyanin, phenylpropanoids, xenbiotics, triterpenes, and coumarins, etc [11]. The enzymes of group A are more specific to the substrates, such as flavonoids, anthocyanidins, and triterpenes. Members of groups B, C, and D are mainly flavonol-7-*O*-glycosyltransferases, benzoic acid glycosyltransferases, and brassinosteroids glycosyltransferases. Most of group F are flavonoid glycosyltransferases [11]. Three *Eu*UGTs were clustered into group O, which are generally associated with the *O*-glycosylation of the hydroxylated isoprenoid side chain of cytokinin [12]. Group O is absent in *A. thaliana* and *L. usitatissimum*, and may have been lost during the evolution of these two species [12]. The members of group R are found in few plants, but *E. ulmoides* has a single UGT occurring in group R, which has been reported to catalyze the *C*-glycosylation and *O*-glycosylation of flavonoids in maize [35] and gallic acid in bamboo [36], respectively.

2.3. Analysis of Conserved Motifs and Prediction of Cis-Acting Elements

All of the *Eu*UGTs' amino acid sequences were submitted to the MEME website for analysis of the conserved motifs. Like the previous analysis [6,37], nine conserved motifs were identified within almost all of the *Eu*UGTs (Figure 2A). However, some *Eu*UGTs may lack one or more motifs. For example, motifs 2 and 3 are missing from three *Eu*UGTs (GWHPAAAL015516, GWHPAAAL024578, and GWHPAAAL024577) in group O (Figure 2A). Motif 5 is considered to separate the UGT sequence into the *N*-terminal and *C*-terminal regions. The *C*-terminal is better conserved than the other terminals. Motif 7—also called the PSPG box—is the UGT-defining consensus sequence and exists across all the putative *Eu*UGTs (Figure 2B). It is almost identical to the PSPG motif within the UGTs of *A. thaliana* and *L. usitatissimum*. The consensus sequence is highly conserved, especially at sites such as 1W, 4Q, 8L, 10H, 16F, 19H, 21GWNS24, 27E, 39P, 43D/E and 44Q. Further analysis showed some differences within the PSPG box in each group of *Eu*UGTs (Figure 3A). In addition to a few positions of motif 7 conserved in most *Eu*UGTs, there are also positions conserved within *Eu*UGT proteins from the same group, such as: 14G in group A; 14G and 28G in group D; 26L, 38W, 42A in group E; and 36V in group L.

Except for the PSPG box, the other eight motifs of the EuUGTs were found to be somewhat conservative (Figure 3B). However, the catalytic functions of these motifs remained unclear until now. By comparing the 3-D structures of several GT-B fold enzymes, it was found that these enzymes showed high structural similarity despite their lower sequence identity [38]. Therefore, it could be speculated that these motifs may play a role in maintaining the stability of the GT-B fold.

The *cis*-acting elements, including the promoter, enhancer, silencer, and various response elements, are important in the regulation of gene transcription. Therefore, the 2000 bp 5′-upstream sequence of each *Eu*UGT gene coding sequence was retrieved and used to predicate the putative *cis*-acting elements on the PlantCARE webserver (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, accessed on 8 December 2020). Overall, 26 various *cis*-acting elements and their locations were predicted (Figure 4). Almost all the *Eu*UGT genes have one or more *cis*-elements, except for GWHPAAAL015755 and GWHPAAAL018171. It is worth noting that some *Eu*UGT genes have multiple tandem elements, such as light-responsive elements (GWHPAAAL014996 and GWHPAAAL018716), low-temperature-responsive elements (GWHPAAAL011200), plant-hormone-responsive elements (GWHPAAAL014206), and wound-responsive elements (GWHPAAAL019177), etc. Therefore, the expression of the *Eu*UGT genes is supposed to be extensively regulated.
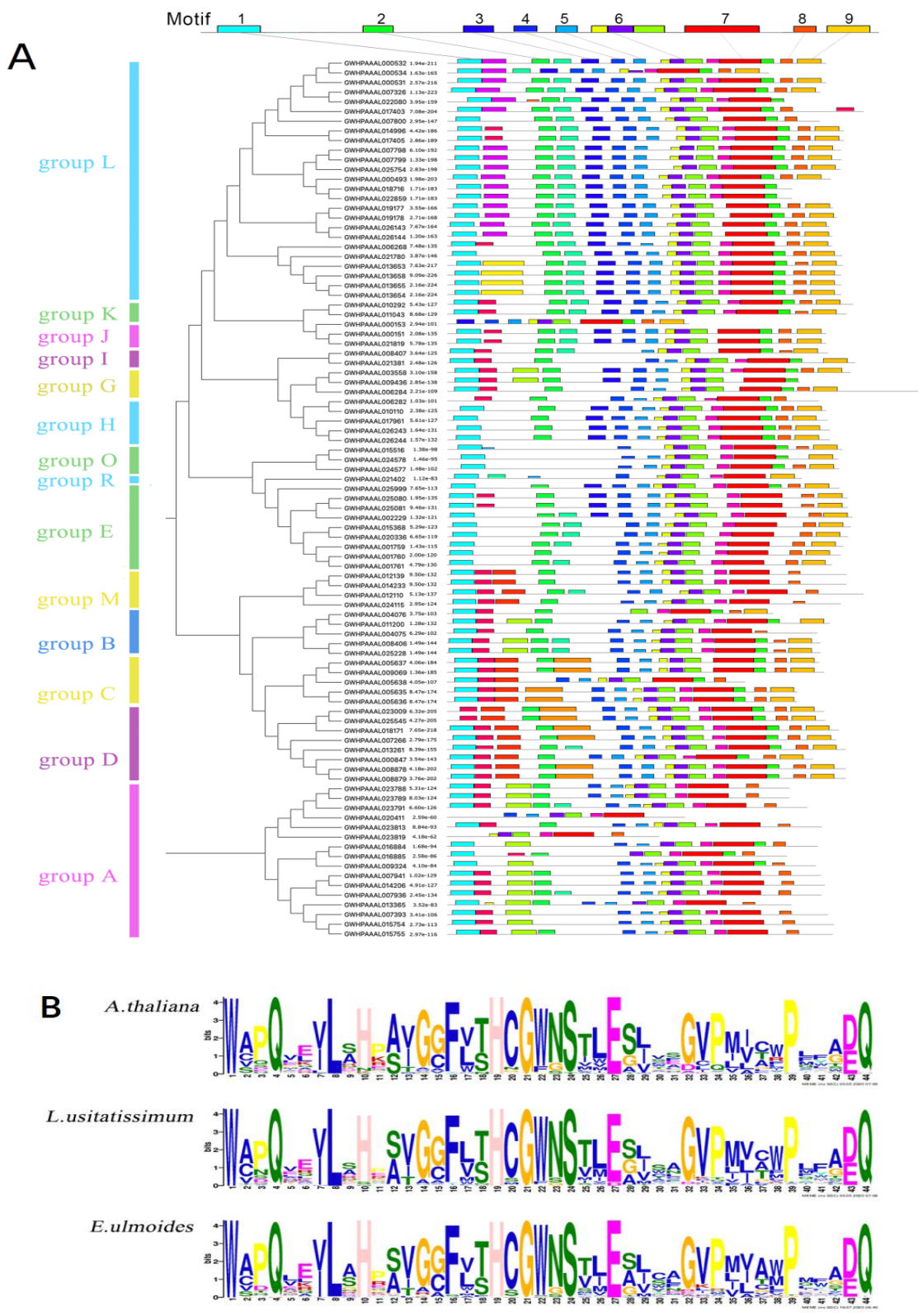
Figure 2. Conservative motifs of *Eu*UGTs (**A**) and the Weblogo of PSPG box (motif 7) within the UGTs in *A. thaliana*, *L. usitatissimum*, and *E. ulmoides* (**B**).
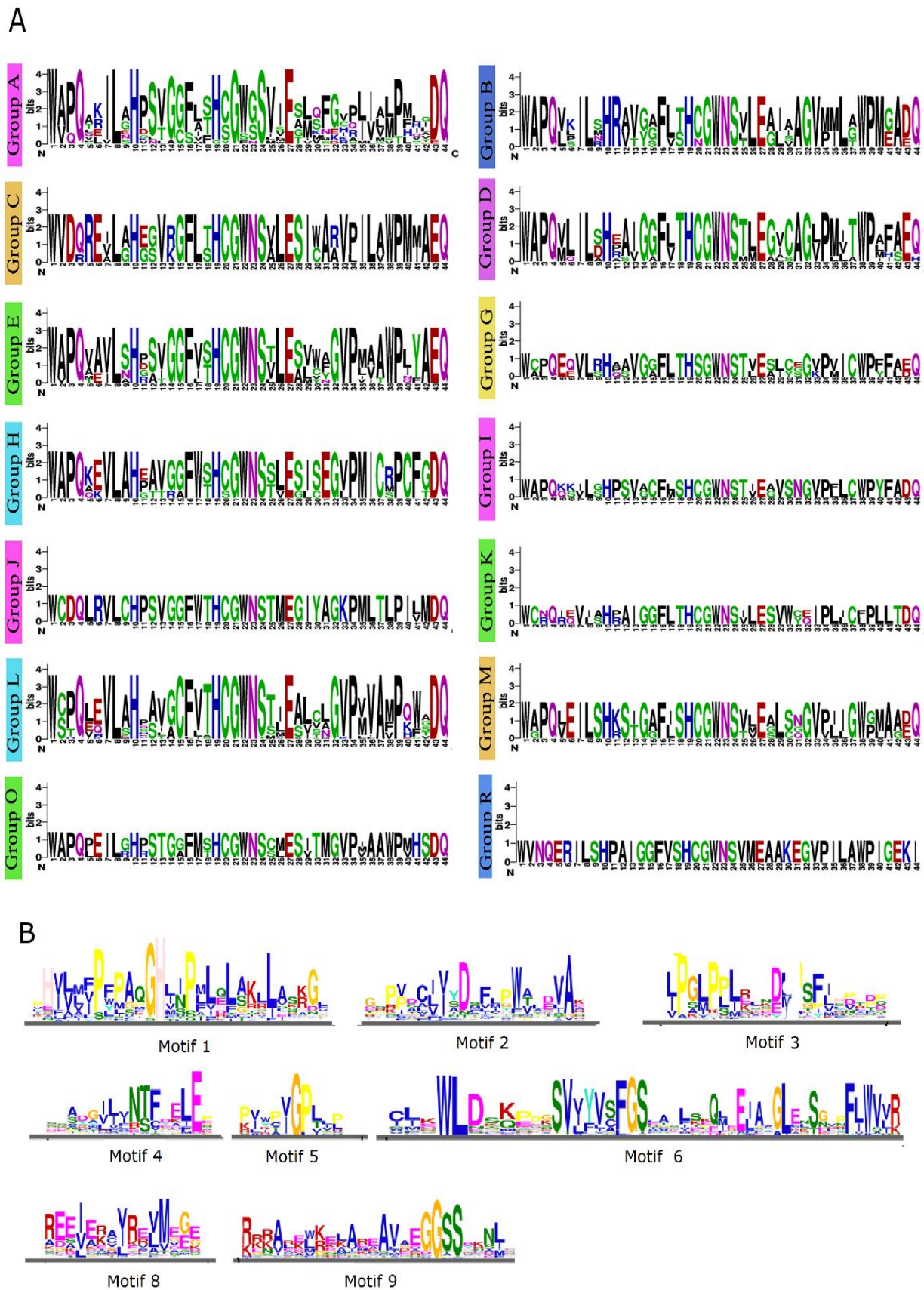
Figure 3. Conservative motifs of *Eu*UGTs. (**A**) The Weblogo of motifs of motif 7 (PSPG box) in each phylogenetic group of *Eu*UGTs; (**B**) the Weblogo of the other motifs in *Eu*UGTs.
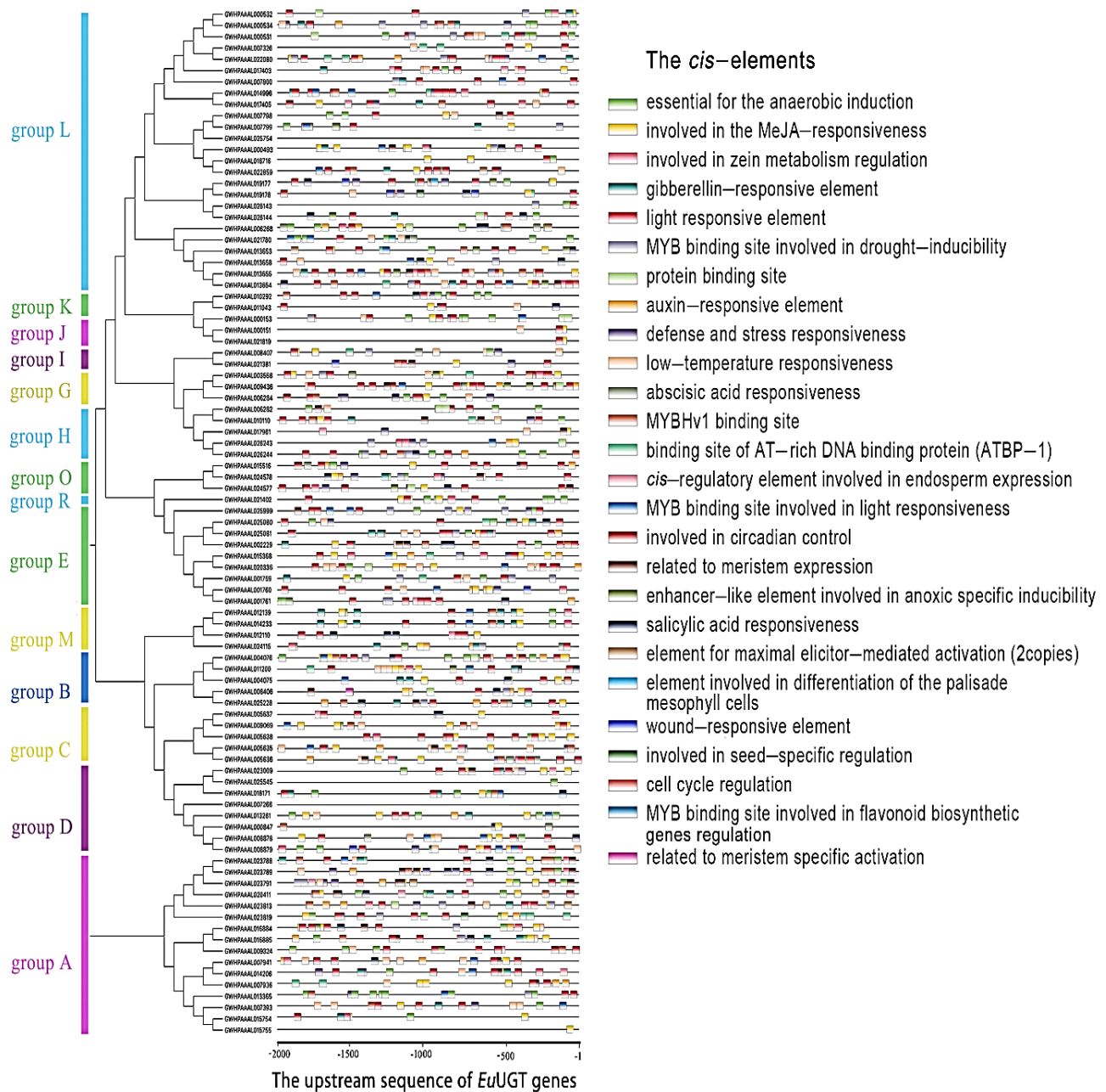
Figure 4. The predicated *cis*-acting elements localized within the 2000-bp upstream of *Eu*UGT genes.

2.4. GO Annotation of EuUGTs

Through GO ontology annotation, the putative *Eu*UGTs were categorized into 45 molecular functions and 93 biological processes (Figure 5A,B, Table S3). Most of the *Eu*UGTs were annotated as flavonoids glucosyltransferase (quercetin 3-*O*-glucosyltransferase and quercetin 7-*O*-glucosyltransferase) and hormonal glycosyltransferase (indole-3-acetate beta-glucosyltransferase, indole-3-butyrate beta-glucosyltransferase, and abscisic acid glucosyltransferase). In addition, GWHPAAAL025999 was annotated as coniferyl alcohol glycosyltransferase, which may be related to the biosynthesis of lignans, such as PDG, a unique and valuable secondary metabolite in *E. ulmoides*. In view of these bioprocesses, most of the *Eu*UGTs were enriched to be involved in the responses to the various stresses, suggesting that the glycoside molecules could participate in the defense of abiotic and biotic stresses.
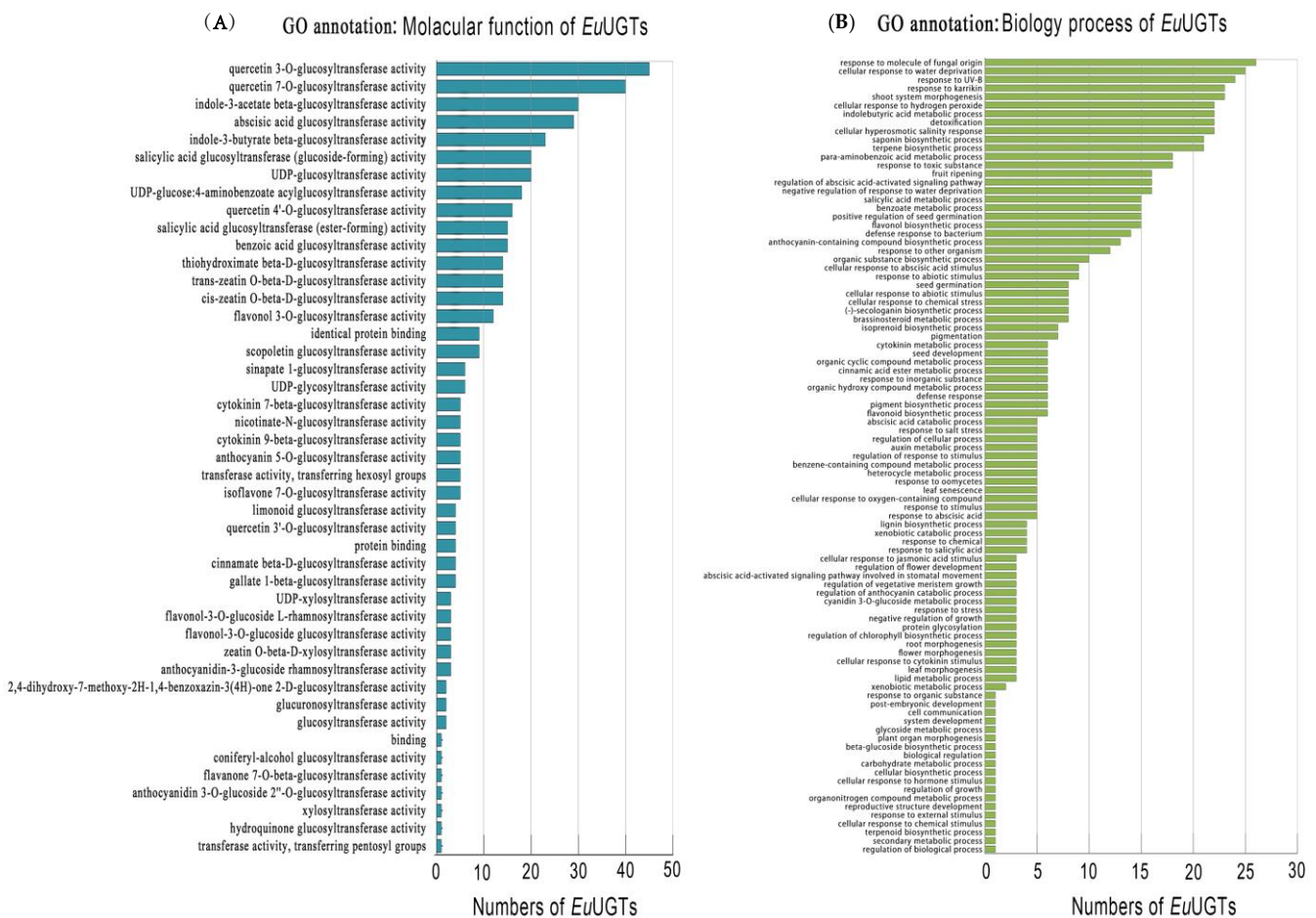
Figure 5. GO annotation of *Eu*UGTs. (**A**) Enrichment of *Eu*UGTs in molecular function; (**B**) enrichment of *Eu*UGTs in biological processes.

2.5. Transcriptome Analysis for Organ-Specific and Time-Specific Gene Expression

The third-party transcriptome data [33] were used to analyze the expression patterns of the *Eu*UGT genes in various organs and in different seasons. The relative transcriptional levels of the *Eu*UGT genes are summarized in Table S4, with TPM values ranging from 0 to 277,336, indicating that most of the *Eu*UGT genes were differentially expressed. Figure 6A shows the expression patterns of *Eu*UGT genes in various organs. Globally, many *Eu*UGT genes exhibited an organ-specific expression pattern, which was highly expressed in one or two organ(s), but lower in others. Based on the expression levels of all of the *Eu*UGT genes, the five organs could be roughly clustered into three groups: Root, bark/stem, and flower/leaf. The root showed a greater difference from the other four organs in terms of *Eu*UGT gene transcription. Some *Eu*UGT genes could be recognized as organ-specific based on their transcriptional level. For example, GWHPAAAL009436 and GWHPAAAL020336 tended to be expressed at higher levels in the root than in other organs; GWHPAAAL023788 was only expressed in the bark and stem. In contrast, GWHPAAAL007266 was highly expressed in the flower, but had extremely low expression in the other four organs. It was observed that all the eight *Eu*UGT genes from group D were expressed primarily in the root and flower. Meanwhile, 6 out of the 16 *Eu*UGT genes belonging to group A had higher expression in the flower than in the other organs. In addition, there were more highly expressed genes in the leaf than in the other four organs, which may be consistent with the fact that plant leaves contain many polysaccharides [39].
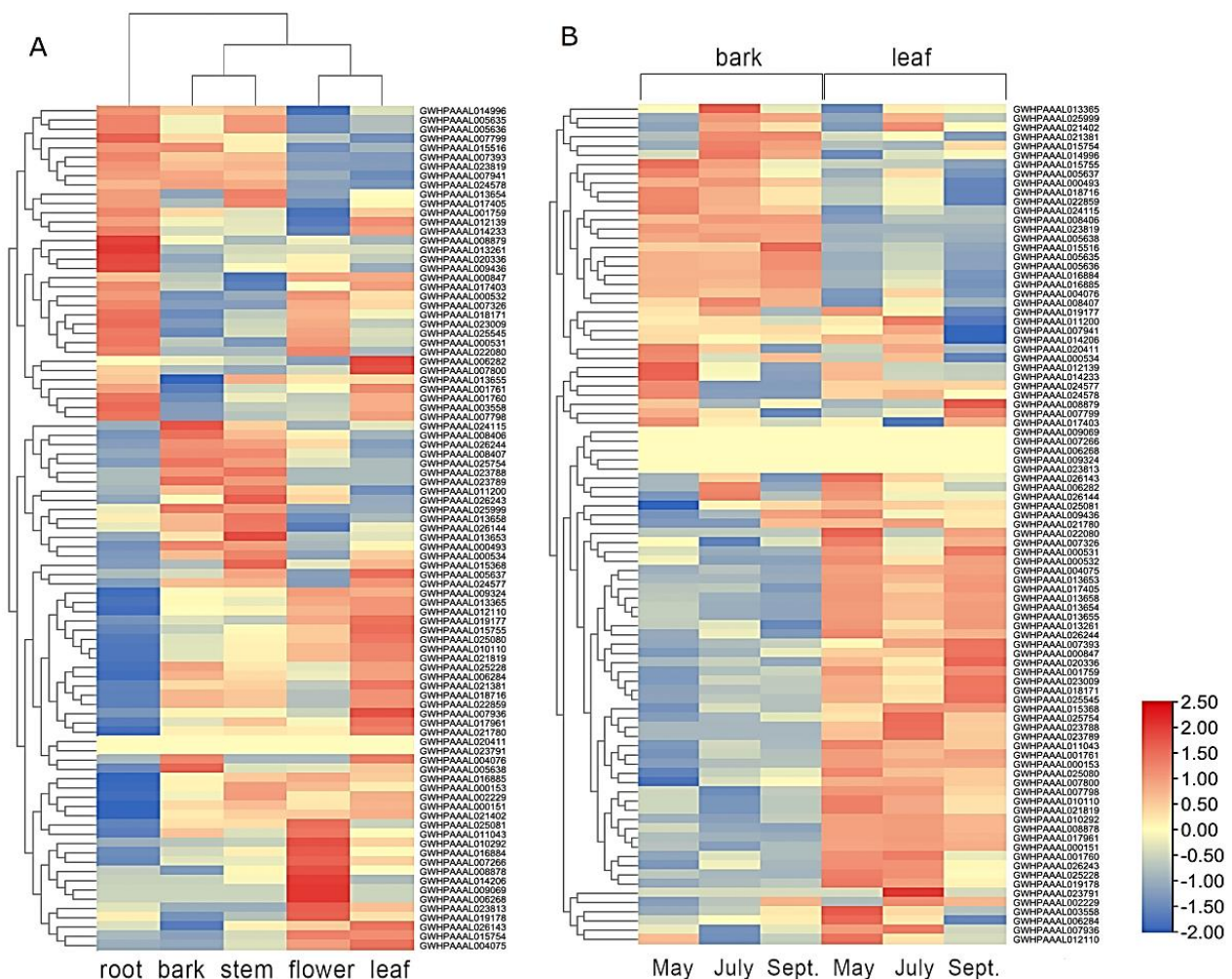
Figure 6. The transcriptome analysis of *Eu*UGT genes. (**A**) Heatmap of *Eu*UGT gene transcription in various organs. (**B**) Heatmap of *Eu*UGT gene transcription over various seasons. The TPM value of each *Eu*UGT gene was calculated by Stringtie and normalized with RowScale method by TBtools, and then used to construct the heatmap. In each row, red color represents higher expression level, blue color shows lower expression.

Furthermore, several *Eu*UGTs, such as GWHPAAAL015368, GWHPAAAL017961, GWHPAAAL000493, GWHPAAAL018716, GWHPAAAL022859, GWHPAAAL025754, GW-HPAAAL025999, and GWHPAAAL07799, had higher expression levels in all five organs, with a TPM value greater than ten thousand, suggesting that these genes might be widely involved in the entire growth and development process.

The contents of many secondary metabolites exhibited a fluctuation over the seasons in *E. ulmoides*. For example, the content of some active ingredients, including PDG, aucubin, and chlorogenic acid, was reported to be higher from May to July than during the other seasons [40,41]. Therefore, we analyzed UGT gene expression in the bark and leaves during various seasons (May, July, and September). A large number of the *Eu*UGT genes did not show obvious fluctuations in transcription over the seasons in bark or leaves (Figure 6B). Nevertheless, a small fraction of the genes exhibited a season-dependent expression pattern. For example, the expression of genes such as GWHPAAAL000847, GWHPAAAL000531, GWHPAAAL020336, GWHPAAAL025545, and GWHPAAAL018171 in the leaf increased significantly in September compared to May. In contrast, some genes' expression levels in the leaf were much higher in May compared to September. A similar situation occurred in the bark. Genes including GWHPAAAL025080, GWHPAAAL025081, GWHPAAAL025999, GWHPAAAL021381, GWHPAAAL007800, and GWHPAAAL015516 had significantly increased expression in the bark during September compared to May.

2.6. RT-qPCR Analysis of the Selected EuUGT Genes

Lignans, generated by the polymerization of three major monolignols (ρ-coumaryl, coniferyl, and sinapyl alcohols), constitute a major class of secondary metabolites [42,43]. In *E. ulmoides*, up to 46 lignans were identified in various organs or tissues. Most of these lignans belong to glycosidic compounds [16]. In order to discover the key UGT genes possibly involved in the glycosylation of lignans in *E. ulmoides*, RT-qPCR was performed to detect the expression level of six selected *Eu*UGTs belonging to groups E and C in root, bark, and leaves, which were sampled from three 18-year-old trees in the harvesting season (in June). Figure 7 shows the relative expression levels of the six genes in the three samples. Among them, GWHPAAAL002229, GWHPAAAL015368, and GWHPAAAL025081 were expressed more efficiently in root than that in the bark and leaves. The expression levels in the root were 5.5, 6.4 and 6.9 times as high as those in bark, and 7.0, 15.3 and 9.1 times as high as those in the leaves, respectively. The transcriptional level of GWHPAAAL005635 was increased by 5.9- and 11.3-fold in bark compared to the root and leaf, respectively. GWHPAAAL001760 had a higher expression level in the bark and root than in the leaves. The expression of GWHPAAAL025999 was noticeably up-regulated in leaf, but its transcripts were observed at a high level in all three organs.
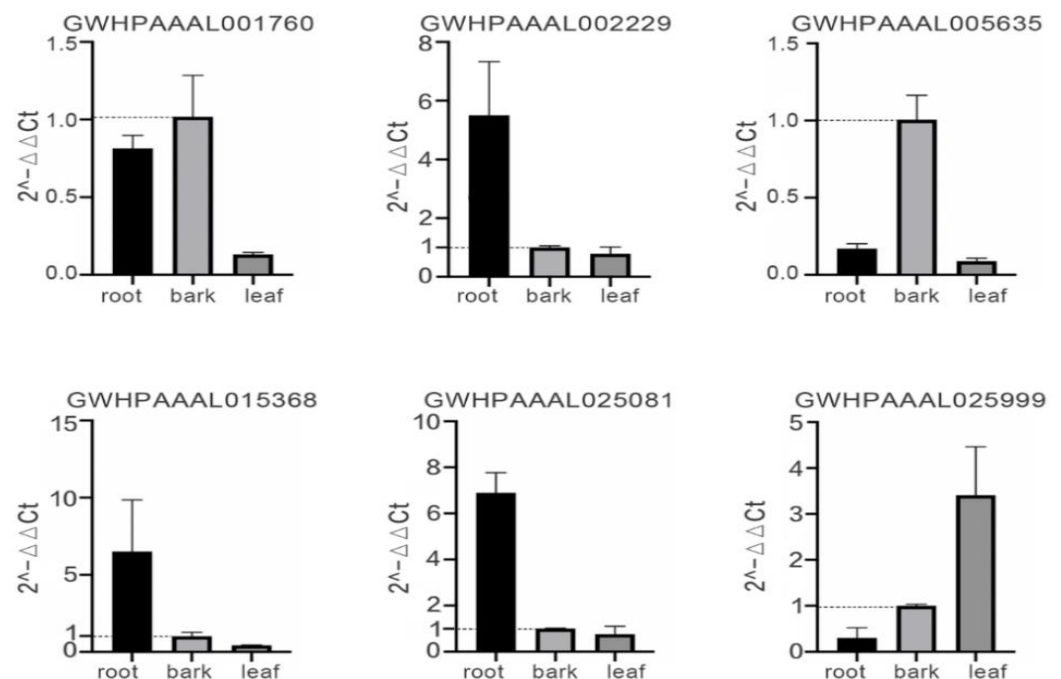

Figure 7. Quantitative real-time PCR analysis of six selected *Eu*UGT genes in root, bark, and leaves of *E. ulmoides*. The relative expression level of each *Eu*UGT gene was normalized to that in the bark. The β-actin was used as the reference gene.

2.7. Construction of Protein–Protein Interaction (PPI) Network of EuUGTs

All 91 *Eu*UGTs were used to construct the PPI network on the String web server (https://string-db.org/, accessed on 9 September 2021) [44]. As a result, a PPI network was obtained (Figure 8), in which nine *Eu*UGTs constituting three nodes (in blue circle) were predicted to be interacted with ten protein partners. All of the proteins included in this network are summarized in Table S5. The UGT79 (GWHPAAAL009324) was predicted to be interacted with 10 protein partners, such as BAN [NAD(P)-binding Rossmann-fold superfamily protein], DFR (dihydroflavonol reductase), FLS1 (flavonol synthase/flavanone 3-hydroxylase), FLS3 (flavonol synthase), TT4 (flavonol synthase), TT5 (chalcone-flavanone isomerase), TT7 (cytochrome P450), F3H (naringenin-2-oxoglutarate 3-dioxygenase), LDOX (leucoanthocyanidin dioxygenase), and UGT75 (anthocyanidin 5-*O*-glucosyltransferase).

UGT91 (GWHPAAAL007941, GWHPAAAL015754 and GWHPAAAL015755) were interacted with five proteins (BAN, FDR, FLS1, TT5, and F3H). UGT90 (GWHPAAAL005635, GWHPAAAL005636, GWHPAAAL005637, GWHPAAAL005638, and GWHPAAAL009069) were only interacted with two proteins (DFR and BAN).
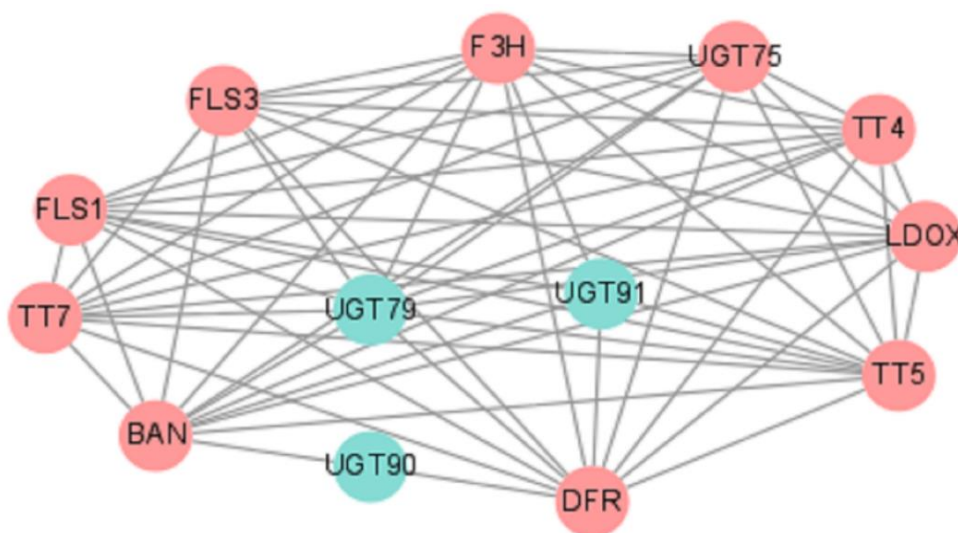

Figure 8. Protein–protein interaction network of *Eu*UGTs with *A. thaliana* as reference predicted by using String web server. The blue circles represent the input *Eu*UGTs; the red circles show the protein partners. BAN, NAD(P)-binding Rossmann-fold superfamily protein; DFR, dihydroflavonol reductase; FLS1, flavonol synthase/flavanone 3-hydroxylase; FLS3, flavonol synthase; TT4, flavonol synthase; TT5, chalcone-flavanone isomerase; TT7, cytochrome P450; F3H, naringenin,2-oxoglutarate 3-dioxygenase; LDOX, leucoanthocyanidin dioxygenase; and UGT75, anthocyanidin 5-*O*-glucosyltransferase.

In view of the catalytic function of the proteins in the network, three possible biological processes were assigned in the biosynthesis of flavonoid, anthocyanin-containing compound, and pigment. For example, a large fraction of the protein partners, such as DFR, FLS1, TT5, TT4, TT7, FLS3, is involved in the biosynthesis of flavonoid; UGT75 and LDOX are included in the biosynthesis of anthocyanin-containing compound, which is downstream of the flavonoid biosynthesis [45].

3. Discussion

E. ulmoides is one of the important medical plants in East Asia. More than 46 lignans and their glycosides were identified in different tissues or organs of *E. ulmoides* [16]. However, no enzyme was found to be involved in their biosynthesis. Therefore, we performed a comprehensive analysis on the UGTs in silico. By searching the draft genome of *E. ulmoides*, 91 UGT genes were identified, which is less than the number in other plants, which in general have more than 100 UGTs. For example, *A. thaliana* contains 123 UGTs [6]; the tree *Quercus robur* has 244 UGTs [46]. In addition, three *Eu*UGTs (GWHPAAAL000153, GWHPAAAL020411, and GWHPAAAL023819) are clearly missing an *N*-terminal. This may be due to the incomplete genome of *E. ulmoides*. The *E. ulmoides* draft genome was assembled previously, with a size of 1.18 Gb which consisted of 29,348 scaffolds with an N50 of less than 1.03 Mb [33]. More than 10 pairs of *Eu*UGT sequences exhibited high similarity (Table S1). The same phenomenon was found in the other plants and is thought to be a result of tandem duplication during long-term evolution [2,6,47]. It may also reflect that the plant UGTs expand in order to produce more secondary metabolites [11,12]. As in other plant species, *Eu*UGT genes are mostly intronless, except for a small fraction of family members containing one to three introns, which may be derived from gain or loss events during their evolution [47].

The *Eu*UGTs were categorized into 14 previously identified groups, such as A to E, G to M, and O and R. Although the amnio acid sequence similarity is low among the UGT groups, the protein structure of the *Eu*UGTs was shown to be highly conserved. Nine motifs could be distinguished among the *Eu*UGTs (Figure 2A). Among these, motif 7 (PSPG box) was previously discussed extensively in an analysis of plant UGTs due to the highly conserved amino acid residues [5,6]. All 91 of the *Eu*UGTs also have a PSPG box with nearly identical conservative amino acid residues (Figure 2B). The function of the PSPG box is to serve as a binding domain of UDP-sugar [5]. Masada et al. (2007) replaced the PSPG box of *Ca*UGT2 with the *Nt*GT1b's, leading to the loss of catalytic activity in the *Ca*UGT2 variant, indicating its important role in catalysis [48]. Further, site-directed mutagenesis of a non-conservative Cys377 in the PSPG box of *Ca*UGT2 to Arg led to the variant losing catalytic activity [48], suggesting that the non-conservative residues in the PSPG box in *Ca*UGT2 may also participate in substrate recognition. As Figure 2A shows, the other eight motifs were found to be conservative to a certain extent, but almost no studies have focused on them. Through *C*-terminal domain swapping, it was found that the affinity of the chimeric enzyme of *At*UGT71C1 and *At*UGT71C2 to the substrate etoposide changed significantly, suggesting that the substrate specificity of UGT may have resided in the *N*-terminal as well as the *C*-terminal domains [49]. Taken together, both the *C*-terminal and *N*-terminal domains of plant UGTs may be involved in substrate recognition, which is an important possibility to explore further in the future.

The *cis*-elements within the 2000 bp upstream sequence of each *Eu*UGT gene were identified. It was shown that up to 26 *cis*-elements were recognized, suggesting that *Eu*UGTs' transcription might be extensively regulated. For example, after being treated with methyl jasmonate, several UGTs in the adventitious roots of *P. ginseng* were up-regulated [13]; in *Brassica* and *Arabidopsis*, changes in the transcriptional level of many UGTs happened in response to various stresses [50]. Our transcriptome analysis also confirmed that the transcription of most *Eu*UGT genes was organ-specific (Figure 6). In addition, it is worth noting that many of the *cis*-elements were enriched to be related to stress responses, such as low temperatures, wounds, and the stresses caused by bacteria and fungi, etc. GO annotation also categorized these *Eu*UGTs into the biological processes of stress responses (Figure 5B). Taken together, the tissue-specific expression of plant UGTs suggests that specific glycosylation may largely take place in a given tissue or organ. For example, several UGT genes were reported to be highly expressed in germinating seeds of *Cicer arietinum* and were localized in the regions of rapidly dividing cells; thus the authors suggested that these tissue-specific expressed UGTs may be involved in cell cycle regulation [51]. Furthermore, the transcriptomic data may help to identify the UGTs potentially involved in the biosynthetic pathways of a given glycoside [13,14,52].

Since many bioactive compounds isolated from *E. ulmoides* belong to lignan and their glycosylated derivatives, such as aucubin, PDG, etc., it is interesting to identify the genes involved in their biosynthesis [25,26,53]. For example, PDG is a kind of natural glycosylated lignan existing in several plant species, such as *E. ulmoides* [16], *L. usitatissimum* [54], *Actinidia arguta* [55], and *Valeriana officinalis* [56]. Previous studies have proposed that the upstream biosynthesis pathway of PDG starts from phenylalanine, then synthesizes cinnamic acid, coumaric acid, coffee acid, ferulic acid, and finally coniferyl alcohol, through the phenylpropanoid pathway [57]. This pathway has been reconstituted in *E. coli*, which could utilize glucose as a substrate to produce coniferyl alcohol [58]. Coniferyl alcohol has been proposed as a precursor molecule to the synthesis of PDG through dimerization to form pinoresinol, which is then glycosylated to produce PDG [59,60]. The pathway of PDG biosynthesis from coniferyl alcohol in plants is not defined and no enzyme has been identified to catalyze the glycosylation of pinoresinol to produce PDG. However, a coniferyl alcohol glucosyltransferase was purified and characterized biochemically from the cambial sap of spruce (*Picea abie* L.) [61]. Two genes, UGT71A18 (found in *Forsythia koreana* [62]) and UGT71C1 (found in *A. thaliana* [63]), were shown to catalyze the glycosylation of pinoresinol to produce pinoresinol monoglucoside. In addition, some other UGTs were

also found to be related to the lignin biosynthesis [52,64–66]. For this purpose, we screened out six *Eu*UGTs for an expression analysis by combining the sequence characteristics and their expression patterns in various organs. GWHPAAAL001760 and GWHPAAAL015368 were clustered into group E with a sequence similarity close to UGT71, and two proteins belonging to this clade in *F. koreana* and *A. thaliana* were reported to catalyze pinoresinol to form pinoresinol monoglucoside [62,63]. In addition, GWHPAAAL025999 was annotated to be coniferyl alcohol glucosyltransferase. The other two *Eu*UGTs (GWHPAAAL002229 and GWHPAAAL025081) were also selected from group E. As shown in our transcriptomic data, another *Eu*UGT (GWHPAAAL005635) showed higher transcription levels in the root and bark, which were previously reported to contain higher PDG content [67,68]. Thus, an RT-qPCR was performed to verify the transcription levels of the selected UGT genes in root, bark, and leaves. It was shown that five selected *Eu*UGT genes, with the exception of GWHPAAAL025999, were transcribed much more efficiently in root or bark than in the leaf, which is consistent with the higher content of PDG and other glycosidic metabolites in the bark and root [67,68]. However, a deviation appears between the qPCR data and transcriptome profiling for some *Eu*UGT genes. For example, the highest expression of the GWHPAAAL005636 gene was observed in bark when using a qPCR analysis but appeared to be in the root when using a transcriptome analysis. These differences may arise from the different sampling locations and trees used in each analysis. Taken together, these results suggest that these *Eu*UGTs could be candidates for future research on the biosynthesis of PDG and other glucosides. The catalytic functions of the selected *Eu*UGTs should be verified experimentally. For this purpose, our laboratory is studying the cloning and expression of the cDNAs and the encoding these *Eu*UGTs.

In addition to the lignan compounds, various organs or tissues of *E. ulmoides* also contain plenty of flavonoids and their glycoside derives, such as quercetin, hyperin, and rutin [16]. Thus, we construct a PPI network with 91 *Eu*UGTs to identify the UGTs that may be involved in biosynthesis of flavonoids and their glycosides. Through the PPI network, a biosynthetic pathway of flavonoids in *E. ulmoides* could be schemed based on the previous proposal [45,69]. As shown in Figure 9, almost all the key enzymes can be deduced in *E. ulmoides* (Table S5) in this pathway. In the PPI network, UGTs constitute three nodes. UGT79 (GWHPAAAL009324) from group A was annotated to encode anthocyanidin 5-*O*-glucosyltransferase, which may participate in the biosynthesis of anthocyanidin glycosides. The other two nodes consist of UGT90 (belonging to group C, containing four *Eu*UGT members: GWHPAAAL005635, GWHPAAAL005636, GWHPAAAL005637, GWHPAAAL005638, GWHPAAAL009069) and UGT91 (belonging to group A, containing three members: GWHPAAAL007941, GWHPAAAL015754, and GWHPAAAL015755) in *E. ulmoides*. All of the above *Eu*UGTs may be involved in the conversion of quercetin or kaempferol into various quercetin or kaempferol glycosides, respectively, such as isoquercitrin, rutin and so on. However, the exact *Eu*UGT cannot be assigned into each glycosylation reaction such as that present in Figure 9, which must be confirmed experimentally.
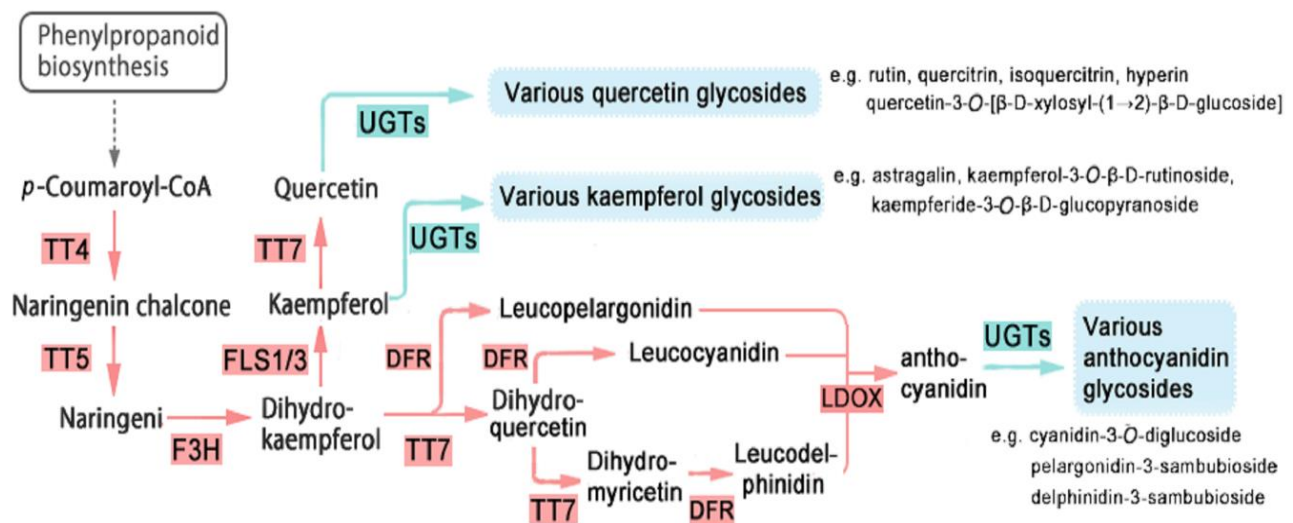
Figure 9. A proposal scheme of biosynthesis of flavonoids and their glycosides in *E. ulmoides*. BAN, NAD(P)-binding Rossmann-fold superfamily protein; DFR, dihydroflavonol reductase; FLS1, flavonol synthase/flavanone 3-hydroxylase; FLS3, flavonol synthase; TT4, flavonol synthase; TT5, chalcone-flavanone isomerase; TT7, cytochrome P450; F3H, naringenin,2-oxoglutarate 3-dioxygenase; LDOX, leucoanthocyanidin dioxygenase; and UGT75, anthocyanidin 5-*O*-glucosyltransferase; UGT, UDP-Glycosyltransferase.

4. Materials and Methods

4.1. Data Resources Used

The draft genome and annotation of *E. ulmoides* [33] were downloaded from the National Genomics Data Center under BioProject accession PRJCA000677 (https://bigd.big.ac.cn/bioproject/browse/PRJCA000677, accessed on 17 March 2020). The transcriptome data of *E. ulmoides* were downloaded from NCBI under BioProject accession PRJNA357336. One hundred and seven UGT sequences of *A. thaliana* (two UGT80s not included) were obtained from Uniprot (https://www.uniprot.org/, accessed on 17 March 2020) and 137 UGT sequences of *Linum usitatissimum* L. were obtained from NCBI (JN088282 to JN088418). In addition, the UGT708C1 (XP_007216617) of *Prunus persica* [70] and the UGTPg36 (AKA44596.1) of *P. ginseng* [71] were used to construct the phylogenetic tree.

4.2. Genome-Wide Identification of Putative UGTs in E. ulmoides

The hidden Markov model UDPGT (PF00201.18) obtained from the Pfam database (http://pfam.xfam.org/, accessed on 5 May 2020) was used to search for the putative UGTs against the *E. ulmoides* draft genome by employing the HMMER-3.3 software with a cut-off *E*-value 1×10^{-5} [72]. Key word searches were also performed through genome annotation using the words 'UGT', 'glycosyltransferase', and 'glucosyltransferase'. All sequences screened by the two above methods were merged and de-duplicated. ClustalW was used to perform multiple sequence alignment and the MEME web server (http://meme-suite.org/tools/meme, accessed on 6 May 2020) [73] was used to find conserved motifs within the candidate *Eu*UGTs. We re-defined the complete sequence of *Eu*UGTs, since parts of genes appear to lack the *N*-terminal or the *C*-terminal when compared to the UGTs of *A. thaliana* and *L. usitatissimum*. The upstream and downstream sequences of putative UGT genes were extracted by Python 3.7 in order to re-annotate them using FGENESH and Softberry Online Services (http://www.softberry.com/, accessed on 1 September 2020) [74]. The transcriptome was assembled with and without the reference genome, using Trinity software [75] in order to help to determine the correct CDS of putative genes. After removing sequences with conservative motifs of less than 3 and without a complete PSPG box, the putative *Eu*UGTs were obtained and further verified by matching the Pfam families on the HMMER website (https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan, accessed on 13 September 2020).

4.3. Prediction of the Cellular Location, Physical and Chemical Properties of EuUGTs

The subcellular localization of putative *Eu*UGTs was analyzed by the CELLO v2.5 online server (http://cello.life.nctu.edu.tw/, accessed on 18 September 2020) [76,77]. The Protparam tool on the ExPASy web server (https://web.expasy.org/protparam/, accessed on 18 September 2020) [78] was used to the calculate sequence length, molecular weight, theoretical pI, instability index, and the grand average of hydropathicity of the *Eu*UGTs.

4.4. Phylogenetic Analysis of UGTs among A. thaliana, L. usitatissimum and E. ulmoides

The amino acid sequences of the selected UGTs, including 107 UGTs of *A. thaliana*, 137 UGTs of *L. usitatissimum*, and 91 putative UGTs of *E. ulmoides*, along with UGT708C1 as a member of phylogenetic group R [70] and UGTPg36 as a member of phylogenetic group O [71], were aligned by ClustalW in MEGAX software [79]. Then, the output file was used to construct a neighbor-joining tree with the p-distance method and pairwise deletion.

4.5. Analysis of Conserved Motifs and Prediction of Cis-Acting Elements

MEME web server [73] was used to search conserved motifs within all the *Eu*UGTs. The 44 residues of the PSPG box at the *C*-terminus were selected to draw a motif logo using the online Weblogo server (https://weblogo.berkeley.edu/logo.cgi, accessed on 18 September 2020) [80]. PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html, accessed on 20 September 2020) was applied to predict the *cis*-acting elements within the 2000 bp upstream of each *Eu*UGT gene [81].

4.6. GO Annotation of EuUGTs

The protein sequences of putative *Eu*UGTs were used to blast against 109 UGTs of *A. thaliana* using Omicsbox software followed with Blast2Go [82] to obtain the molecular function and biological process of each *Eu*UGT.

4.7. Transcriptome Analysis for Organ-Specific and Season-Specific Gene Expression

The RNA-seq datasets generated from different organs and seasons were downloaded from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra/, accessed on 12 July 2020) [33]. The information on plant samples and the RNA-seq are summarized in Table S6. The expression level of each *Eu*UGT gene were calculated in TPM value by using the Hisat2 [83], Samtools, and Stringtie [84] software, during which 91 *Eu*UGT gene sequences were input as references. The heatmaps of *Eu*UGT gene expression patterns were constructed by TBtools [85].

4.8. Quantitative Real-Time PCR Analysis of the Selected EuUGT Genes

In June 2020, three plant samples of *E. ulmoides*, including bark, leaves, and root, were collected from three 18-year-old trees at a local *E. ulmoides* forest farm (103° E, 30° N) in Chengdu, China. A piece of bark approximately 5×5 cm^2 in size was peeled from each individual trunk at approximately 0.5 m above the ground. The roots were collected from the fibrous roots with a diameter of approximately 5 mm. Three mature leaves were sampled on an individual tree. All of the plant materials were immediately immersed in liquid nitrogen after sampling. The plant materials were smashed in liquid nitrogen, and then used to isolate the complete RNAs using the CTAB protocol [86]. Reverse transcription and the real-time PCR were performed using the PrimeScript RT kit and TB Green Premix Ex Taq II (Takara Bio Co. Dalian, China) following the manufacturer's guides, respectively. Six *Eu*UGT genes were selected to confirm the organ-specific expression, with β-actin as the reference gene. The primers are listed in Table S7. The qPCR reaction was performed with the following parameters: An initial denaturation for 30 s at 95 °C; then, 45 cycles, including denaturation, for 5 s at 95 °C; annealing for 30 s at 55 °C; and extension for 30 s at 72 °C. The result was calculated with the following formula [87]:

$$2^{-\Delta\Delta Ct} = 2^{-(\Delta Ct(\text{root/leaf}) - \Delta Ct(\text{bark}))}; \Delta Ct(\text{root/bark/leaf}) = Ct(\text{root/bark/leaf}) - Ct(\text{actin})$$

4.9. Construction of Protein–Protein Interaction (PPI) Network of EuUGTs

In order to exploit the biological functions of *Eu*UGTs, a PPI network was constructed. Briefly, 91 sequences of *Eu*UGTs were input into the String web server with *A. thaliana* as the reference (https://string-db.org/, accessed on 9 September 2021) [44] to predict the PPI network. The generated result was imported into Cytoscape [88] for plotting.

5. Conclusions

An increasing number of studies have shown that UGTs play an important role in various biological processes in organisms. Therefore, the identification and in silico analysis of UGT gene families in different species have attracted many researchers. Up to now, almost no studies have been performed on the UGT family of *E. ulmoides*, an important pharmacological woody plant. Here, we identified 91 *Eu*UGTs in *E. ulmoides* through genome-wide analysis, which could be clustered into 14 out of the 18 previously described phylogenetic groups (A to R). However, no *Eu*UGT was categorized into groups F, N, P, or Q. Group L was the largest group of *Eu*UGTs, which has a wide variety of substrates. Through transcriptome analysis, differentially expressed patterns were observed for most of the *Eu*UGT genes, which tended to be expressed at higher levels in one or two organs. The organ-specific high expression of the *Eu*UGT genes may be due to the specific metabolism requirements in the organs or in response to various stresses. Through the in silico analysis of the protein–protein interaction network of *Eu*UGTs, a set of proteins including some UGTs were identified, perhaps to constitute a biosynthetic pathway of flavonoids and their glycosides in *E. ulmoides*. Combining multiple types of information from the phylogenetic analysis, the structural characteristics, and the function predictions was helpful to our exploration of the UGTs of interest. Our results could lay the foundation for further studies on UGTs and the biosynthesis of secondary metabolites of *E. ulmoides*.

References

1. Campbell, J.A.; Davies, G.J.; Bulone, V.; Henrissat, B. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* **1997**, *326*, 929–939. [CrossRef] [PubMed]
2. Coutinho, P.M.; Deleury, E.; Davies, G.J.; Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **2003**, *328*, 307–317. [CrossRef]
3. Rahimi, S.; Kim, J.; Mijakovic, I.; Jung, K.H.; Choi, G.; Kim, S.C.; Kim, Y.J. Triterpenoid-biosynthetic UDP-glycosyltransferases from plants. *Biotechnol. Adv.* **2019**, *37*, 107394. [CrossRef] [PubMed]
4. le Roy, J.; Huss, B.; Creach, A.; Hawkins, S.; Neutelings, G. Glycosylation is a major regulator of phenylpropanoid availability and biological activity in plants. Front. *Plant Sci.* **2016**, *7*, 735. [CrossRef]

5. Mackenzie, P.I.; Owens, I.S.; Burchell, B.; Bock, K.W.; Bairoch, A.; Bélanger, A.; Fournel-Gigleux, S.; Green, M.; Hum, D.W.; Iyanagi, T.; et al. The UDP glycosyltransferase gene superfamily: Recommended nomenclature update based on evolutionary divergence. *Pharmacogenet* **1997**, *7*, 255–269. [CrossRef]
6. Li, Y.; Baldauf, S.; Lim, E.K.; Bowles, D.J. Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. *J. Biol. Chem.* **2001**, *276*, 4338–4343. [CrossRef]
7. Li, Y.J.; Li, P.; Wang, Y.; Dong, R.R.; Yu, H.M.; Hou, B.K. Genome-wide identification and phylogenetic analysis of family-1 UDP glycosyltransferase in maize (*Zea mays*). *Planta* **2014**, *239*, 1265–1279. [CrossRef] [PubMed]
8. He, Y.; Ahmad, D.; Zhang, X.; Zhang, Y.; Wu, L.; Jiang, P.; Ma, H. Genome-wide analysis of family-1 UDP glycosyltransferases (UGT) and identification of UGT genes for FHB resistance in wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **2018**, *18*, 67. [CrossRef] [PubMed]
9. Yonekura-Sakakibara, K.; Hanada, K. An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.* **2011**, *66*, 182–193. [CrossRef]
10. Barvkar, V.T.; Pardeshi, V.C.; Kale, S.M.; Kadoo, N.Y.; Gupta, V.S. Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in *Linum usitatissimum* identified genes with varies expression patterns. *BMC Genom.* **2012**, *13*, 175. [CrossRef]
11. Wilson, A.E.; Tian, L. Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism. *Plant J.* **2019**, *100*, 1273–1288. [CrossRef] [PubMed]
12. Caputi, L.; Malnoy, M.; Goremykin, V.; Nikiforova, S.; Martens, S. A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.* **2012**, *69*, 1030–1042. [CrossRef] [PubMed]
13. Kang, K.B.; Jayakodi, M.; Lee, Y.S.; Nguyen, V.B.; Park, H.S.; Koo, H.J.; Choi, I.Y.; Kim, D.H.; Chung, Y.J.; Ryu, B.; et al. Identification of candidate UDP-glycosyltransferases involved in protopanaxadiol-type ginsenoside biosynthesis in *Panax ginseng*. *Sci. Rep.* **2018**, *8*, 11744. [CrossRef] [PubMed]
14. Wen, L.; Yun, X.; Zheng, X.; Xu, H.; Zhan, R.; Chen, W.; Xu, Y.; Chen, Y.; Zhang, J. Transcriptomic comparison reveals candidate genes for triterpenoid biosynthesis in two closely related *Ilex* Species. *Front. Plant Sci.* **2017**, *8*, 634. [CrossRef] [PubMed]
15. Cui, L.; Yao, S.; Dai, X.; Yin, Q.; Liu, Y.; Jiang, X.; Wu, Y.; Qian, Y.; Pang, Y.; Gao, L.; et al. Identification of UDP-glycosyltransferases involved in the biosynthesis of astringent taste compounds in tea (*Camellia sinensis*). *J. Exp. Bot.* **2016**, *8*, 2285–2297. [CrossRef]
16. Wang, C.Y.; Tang, L.; He, J.W.; Li, J.; Wang, Y.Z. Ethnobotany, phytochemistry and pharmacological properties of *Eucommia ulmoides*: A review. *Am. J. Chin. Med.* **2019**, *47*, 259–300. [CrossRef]
17. Wang, J.Y.; Chen, X.J.; Zhang, L.; Pan, Y.Y.; Gu, Z.X.; Yuan, Y. Anti-inflammatory effects of *Eucommia ulmoides* Oliv.: Male flower extract on lipopolysaccharide-induced inflammation. *Chin. Med. J.* **2019**, *132*, 319–328. [CrossRef]
18. Lang, C.; Liu, Z.; Taylor, H.W.; Baker, D.G. Effect of *Eucommia ulmoides* on systolic blood pressure in the spontaneous hypertensive rat. *Am. J. Chin. Med.* **2005**, *33*, 215–230. [CrossRef]
19. Wu, D.; Yu, D.; Zhang, Y.; Dong, J.; Li, D.; Wang, D. Metabolite profiles, bioactivity, and HPLC fingerprint of different varieties of *Eucommia ulmoides* Oliv.: Towards the utilization of medicinal and commercial Chinese endemic tree. *Molecules* **2018**, *23*, 1898. [CrossRef]
20. Li, Y.; Pan, H.; Zhang, X.; Wang, H.; Liu, S.; Zhang, H.; Qian, H.; Wang, L.; Ying, H. Geniposide improves glucose homeostasis via regulating foxo1/pdk4 in skeletal muscle. *J. Agric. Food. Chem.* **2019**, *67*, 4483–4492. [CrossRef]
21. Ma, Z.G.; Kong, C.Y.; Song, P.; Zhang, X.; Yuan, Y.P.; Tang, Q.Z. Geniposide protects against obesity-related cardiac injury through Ampkalpha- and Sirt1-dependent mechanisms. *Oxid. Med. Cell Longev.* **2018**, *2018*, 6053727. [CrossRef]
22. Zhou, Y.; Li, P.; Duan, J.X.; Liu, T.; Guan, X.X.; Mei, W.X.; Liu, Y.P.; Sun, G.Y.; Wan, L.; Zhong, W.J.; et al. Aucubin alleviates bleomycin-induced pulmonary fibrosis in a mouse model. *Inflammation* **2017**, *40*, 2062–2073. [CrossRef]
23. Chang, I.M. Liver-protective activities of aucubin derived from traditional oriental medicine. *Res. Comm. Mol. Pathol. Pharmacol.* **1998**, *102*, 189–204. [CrossRef]
24. Yao, J.; Zou, Z.; Wang, X.; Ji, X.; Yang, J. Pinoresinol diglucoside alleviates oxLDL-induced dysfunction in human umbilical vein endothelial cells. *Evid. Based Complement. Alternat. Med.* **2016**, *2016*, 3124519. [CrossRef]
25. Li, Y.J.; Dong, X.; Zhao, D.G. Cloning and sequence analysis of glucosyltransferase gene EuCGT1 in *Eucommia ulmoides* Olive. *J. Mount. Agri. Biol.* **2014**, *33*, 27–35. [CrossRef]
26. Leng, Y.; Liu, S.H.; Zhao, Y.C.; Zhao, D.G. Prokaryotic expression and protein purification of EUGT2 gene from *Eucommia ulmoides*. *Mol. Plant Breed.* **2020**, *18*, 3240–3245. [CrossRef]
27. Baruah, P.M.; Krishnatreya, D.B.; Bordoloi, K.S.; Gill, S.S.; Agarwala, N. Genome wide identification and characterization of abiotic stress response lncRNA in *Capsicum annuum*. *Plant Physiol. Biochem.* **2021**, *162*, 221–236. [CrossRef]
28. Kumar, M.; Kherawat, B.S.; Dey, P.; Saha, D.; Singh, A.; Bhatia, S.K.; Ghodake, G.S.; Kadam, A.A.; Kim, H.-U.; Manorama; et al. Genome-wide identification and characterization of PIN-FORMED (PIN) gene family reveals role in developmental and various stress conditions in *Triticum aestivum* L. *Int. J. Mol. Sci.* **2021**, *22*, 7396. [CrossRef]
29. Kesawat, M.S.; Kherawat, B.S.; Singh, A.; Dey, P.; Kabi, M.; Debnath, D.; Saha, D.; Khandual, A.; Manorama, S.R.; Ali, A.; et al. Genome-wide identification and characterization of the brassinazole-resistant (BZR) gene family and its expression in the various developmental stage and stress conditions in wheat (*Triticum aestivum* L.). *Int. J. Mol. Sci.* **2021**, *22*, 8743. [CrossRef] [PubMed]
30. Wang, M.; Chen, B.; Zhou, W.; Xie, L.; Wang, L.; Zhang, Y.Z.Q. Genome-wide identification and expression analysis of the AT-hook Motif Nuclear Localized gene family in soybean. *BMC Genom.* **2021**, *22*, 361. [CrossRef]

31. Jiang, M.; Chen, H.; Liu, J.; Du, Q.; Lu, S.; Liu, C. Genome-wide identification and functional characterization of natural antisense transcripts in *Salvia miltiorrhiza*. *Sci. Rep.* **2021**, *11*, 4769. [CrossRef] [PubMed]

32. Tong, T.; Fang, Y.-X.; Zhang, Z.; Zheng, J.; Zhang, X.; Li, J.; Niu, C.; Xue, D.; Zhang, X. Genome-wide identification and expression pattern analysis of the *KCS* gene family in barley. *Plant Growth Reg.* **2021**, *93*, 89–103. [CrossRef]

33. Wuyun, T.; Wang, L.; Liu, H.; Wang, X.; Zhang, L.; Bennetzen, J.L.; Li, T.; Yang, L.; Liu, P.; Du, L.; et al. The hardy rubber tree genome provides insights into the evolution of polyisoprene biosynthesis. *Mol. Plant.* **2018**, *11*, 429–442. [CrossRef]

34. Velasco, R.; Zharkikh, A.; Affourtit, J.; Dhingra, A.; Cestaro, A.; Kalyanaraman, A.; Fontana, P.; Bhatnagar, S.K.; Troggio, M.; Pruss, D.; et al. The genome of the domesticated apple (*Malus x domestica* borkh). *Nat. Genet.* **2010**, *42*, 833–839. [CrossRef] [PubMed]

35. Ferreyra, M.L.F.; Rodriguez, E.; Casas, M.I.; Labadie, G.; Grotewold, E.; Casati, P. Identification of a bifunctional maize *C*- and *O*-glucosyltransferase. *J. Biol. Chem.* **2013**, *288*, 31678–31688. [CrossRef]

36. Sun, Y.; Chen, Z.; Yang, J.; Mutanda, I.; Li, S.; Zhang, Q.; Zhang, Y.; Zhang, Y.; Wang, Y. Pathway-specific enzymes from bamboo and crop leaves biosynthesize anti-nociceptive c-glycosylated flavones. *Comm. Biol.* **2020**, *3*, 110. [CrossRef]

37. Ross, J.; Li, Y.; Lim, E.K. Higher plant glycosyltransferases. *Genome Biol.* **2001**, *2*, reviews3004.1–3004.6. [CrossRef]

38. Shao, H.; He, X.; Achnine, L.; Blount, J.W.; Dixon, R.A.; Wang, X. Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell* **2005**, *17*, 3141–3154. [CrossRef]

39. Xu, J.; Hou, H.; Hu, J.; Liu, B. Optimized microwave extraction, characterization and antioxidant capacity of biological polysaccharides from *Eucommia ulmoides* Oliver leaf. *Sci. Rep.* **2018**, *8*, 6561. [CrossRef]

40. Lan, J.Y.; Liu, Z.; Zhang, Q.; Pei, Z.C.; Li, B.H. A study on growth characteristics of *Eucommia ulmoides* Oliver in different areas and the dynamic changes of main effective components in leaves. *J. Agric. Univ. Hebei* **2019**, *42*, 51–56. [CrossRef]

41. Wang, L.N.; Li, W.; Qin, J.P.; Yang, M.H. Study on dynamic variation of main contents in different parts of *Eucommia ulmoides*. *Chin. Pharm.* **2009**, *18*, 29–31. [CrossRef]

42. Guerriero, G.; Berni, R.; Munoz-Sanchez, J.A.; Apone, F.; Abdel-Salam, E.M.; Qahtan, A.A.; Alatar, A.A.; Cantini, C.; Cai, G.; Hausman, J.F.; et al. Production of plant secondary metabolites: Examples, tips and suggestions for biotechnologists. *Genes* **2018**, *9*, 309. [CrossRef] [PubMed]

43. Satake, H.; Koyama, T.; Bahabadi, S.E.; Matsumoto, E.; Ono, E.; Murata, J. Essences in metabolic engineering of lignan biosynthesis. *Metabolites* **2015**, *5*, 270–290. [CrossRef] [PubMed]

44. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **2021**, *49*, D605–D612. [CrossRef]

45. Koes, R.; Verweij, W.; Quattrocchio, F. Flavonoids: A colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* **2005**, *10*, 236–242. [CrossRef]

46. Zhang, J.; Li, L.-M.; Cheng, W.-W.; Song, X.; Long, Y.-H.; Xing, Z.-B. Genome-wide identification and expression analysis of glycosyltransferase gene family 1 in *Quercus robur* L. *J. Appl. Genet.* **2021**. [CrossRef]

47. Krishnamurthy, P.; Tsukamoto, C.; Ishimoto, M. Reconstruction of the evolutionary histories of UGT gene superfamily in legumes clarifies the functional divergence of duplicates in specialized metabolism. *Int. J. Mol. Sci.* **2020**, *21*, 1855. [CrossRef] [PubMed]

48. Masada, S.; Terasaka, K.; Mizukami, H. A single amino acid in the PSPG-box plays an important role in the catalytic function of CaUGT2 (Curcumin glucosyltransferase), a group D family 1 glucosyltransferase from *Catharanthus roseus*. *FEBS Lett.* **2007**, *581*, 2605–2610. [CrossRef]

49. Hansen, E.H.; Osmani, S.A.; Kristensen, C.; Møller, B.L.; Hansen, J. Substrate specificities of family 1 UGTs gained by domain swapping. *Phytochemistry* **2009**, *70*, 473–482. [CrossRef]

50. Rehman, H.M.; Nawaz, M.A.; Shah, Z.H.; Ludwig-Muller, J.; Chung, G.; Ahmad, M.O.; Yang, S.H.; Lee, S.I. Comparative genomic and transcriptomic analysis of family-1 UDP glycosyltransferase in three *Brassica* species and *Arabidopsis* indicates stress-responsive. *Sci. Rep.* **2018**, *8*, 1875. [CrossRef]

51. Sharma, R.; Rawat, V.; Suresh, C.G. Genome-wide identification and tissue-specific expression analysis of UDP-glycosyltransferases genes confirm their abundance in *Cicer arietinum* (Chickpea) genome. *PLoS ONE* **2014**, *9*, e109715. [CrossRef] [PubMed]

52. Cheng, X.; Muhammad, A.; Li, G.; Zhang, J.; Cheng, J.; Qiu, J.; Jiang, T.; Jin, Q.; Cai, Y.; Lin, Y. Family-1 UDP glycosyltransferase in pear (*Pyrus bretschneideri*): Molecular identification, phylogenomic characterization and expression profiling during stone cell formation. *Mol. Biol. Rep.* **2019**, *46*, 2153–2175. [CrossRef] [PubMed]

53. Liu, Y.-C.; Peng, X.-X.; Lu, Y.-B.; Wu, X.-X.; Chen, L.-W.; Feng, H. Genome-wide association study reveals the genes associated with the leaf inclusion contents in Chinese medical tree *Eucommia ulmoides*. *Biosci. Biotech. Biochem.* **2021**, *85*, 233–241. [CrossRef]

54. Beejmohun, V.; Fliniaux, O.; Hano, C.; Pilard, S.; Grand, E.; Lesur, D.; Cailleu, D.; Lamblin, F.; Laine, E.; Kovensky, J.; et al. Coniferin dimerisation in lignan biosynthesis in flax cells. *Phytochemistry* **2007**, *68*, 2744–2752. [CrossRef] [PubMed]

55. Kwon, D.; Kim, G.D.; Kang, W.; Park, J.E.; Kim, S.H.; Choe, E.; Kim, J.I.; Auh, J.H. Pinoresinol diglucoside is screened as a putative alpha-glucosidase inhibiting compound in *Actinidia arguta* leaves. *J. Korean Soc. Appl. Biol. Chem.* **2014**, *57*, 473–479. [CrossRef]

56. Do, K.H.; Choi, Y.W.; Kim, E.K.; Yun, S.J.; Kim, M.S.; Lee, S.Y.; Ha, J.M.; Kim, J.H.; Kim, C.D.; Son, B.G.; et al. Pinoresinol-4,4′-di-o-beta-d-glucoside from *Valeriana officinalis* root stimulates calcium mobilization and chemotactic migration of mouse embryo fibroblasts. *Phytomedicine* **2009**, *16*, 530–537. [CrossRef] [PubMed]

57. Chen, Z.; Sun, X.; Li, Y.; Yan, Y.; Yuan, Q. Metabolic engineering of *Escherichia coli* for microbial synthesis of monolignols. *Metab. Eng.* **2017**, *39*, 102–109. [CrossRef] [PubMed]

58. Davin, L.B.; Lewis, N.G. An historical perspective on lignan biosynthesis: Monolignol, allylpenol and hydroxycinnamic acid coupling and downstream metabolism. *Phytochemistry* **2003**, *2*, 257–288. [CrossRef]

59. Satake, H.; Ono, E.; Murata, J. Recent advances in the metabolic engineering of lignan biosynthesis pathways for the production of transgenic plant-based foods and supplements. *J. Agric. Food Chem.* **2013**, *61*, 11721–11729. [CrossRef]

60. Zhang, Y.; Shi, J.; Ni, Y.; Liu, Y.; Gao, Z.; Zhao, X.; Gao, Z. Tracing the mass flow from glucose and phenylalanine to pinoresinol and its glycosides in Phomopsis sp. XP-8 using stable isotope assisted TOF-MS. *Sci. Rep.* **2019**, *9*, 18495. [CrossRef] [PubMed]

61. Schmid, G.; Grisebach, H. Enzymic synthesis of lignin precursors. Purification and properties of UDPglucose: Coniferyl-alcohol glucosyltransferase from cambial sap of spruce (*Picea abies* L.). *Eur. J. Biochem.* **1982**, *123*, 363–370. [CrossRef] [PubMed]

62. Ono, E.; Kim, H.J.; Murata, J. Molecular and functional characterization of novel furofuran class lignan glucosyltransferases from *Forsythia*. *Plant Biotech.* **2010**, *27*, 317–324. [CrossRef]

63. Okazawa, A.; Kusunose, T.; Ono, E.; Kim, H.J.; Muranaka, T. Glucosyltransferase activity of *Arabidopsis* UGT71c1 towards pinoresinol and lariciresinol. *Plant Biotechnol.* **2015**, *5*, 561–566. [CrossRef]

64. Wang, Y.W.; Wang, W.C.; Jin, S.H.; Hou, B.K. Over-expression of a putative poplar glycosyltransferase gene, PtGT1, in tobacco increases lignin content and causes early flowering. *J. Exp. Bot.* **2012**, *63*, 2799–2808. [CrossRef]

65. Lin, J.S.; Huang, X.X.; Li, Q.; Cao, Y.; Bao, Y.; Li, X.F.M.Y.J.; Fu, C.; Hou, B.K. UDP-glycosyltransferase 72B1 catalyzes the glucose conjugation of monolignols and is essential for the normal cell wall lignification in *Arabidopsis thaliana*. *Plant J.* **2016**, *88*, 26–42. [CrossRef]

66. Liu, C.; Guo, F.F.; Xiao, J.P.; Wei, J.Y.; Tang, L.Y.; Yang, H.J. Research advances in chemical constituents and pharmacological activities of different parts of *Eucommia ulmoides*. *Chin. J. Chin. Mater. Med.* **2020**, *3*, 497–512. [CrossRef]

67. Yan, Y.; Zhao, H.; Zou, L.S.; Liu, X.H.; Chai, C.; Wang, S.N.; Jiao, Y.J. Difference of chemical constituents in different medicinal parts of *Eucommia ulmoides* by LC-Triple TOF MS/MS. *J. Chin. Mass Spectr. Soc.* **2018**, *39*, 101–111. [CrossRef]

68. Wu, B.; Liu, X.; Zhang, B. Genome-wide characterization, evolution and expression profiling of UDP-glycosyltransferase family in pomelo (*Citrus grandis*) fruit. *BMC Plant Biol.* **2020**, *20*, 459. [CrossRef]

69. Lewis, D.R.; Ramirez, M.V.; Miller, N.D.; Vallabhaneni, P.; Ray, W.K.; Winkel, R.F.H.B.S.J.; Muday, G.K. Auxin and ethylene induce flavonol accumulation through distinct transcriptional networks. *Plant Physiol.* **2011**, *156*, 144–164. [CrossRef]

70. Wu, B.; Gao, L.; Gao, J.; Xu, Y.; Liu, H.; Cao, X.; Zhang, B.; Chen, K. Genome-wide identification, expression patterns, and functional analysis of UDP glycosyltransferase family in peach (*Prunus persica* L. Batsch). *Front. Plant Sci.* **2017**, *8*, 389. [CrossRef]

71. Wang, P.; Wei, Y.; Fan, Y.; Liu, Q.; Wei, W.; Yang, C.; Zhang, L.; Zhao, G.; Yue, J.; Yan, X.; et al. Production of bioactive ginsenosides RH2 and RG3 by metabolically engineered yeasts. *Metab. Eng.* **2015**, *29*, 97–105. [CrossRef]

72. Potter, S.C.; Luciani, A.; Eddy, S.R.; Park, Y.; Lopez, R.; Finn, R.D. Hmmer web server: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W200–W204. [CrossRef]

73. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef] [PubMed]

74. Solovyev, V.; Kosarev, P.; Seledsov, I.; Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **2006**, *7* (Suppl. 1), S10.1–S10.12. [CrossRef]

75. 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. Biotech.* **2011**, *29*, 644–652. [CrossRef] [PubMed]

76. Yu, C.S.; Lin, C.J.; Hwang, J.K. Predicting subcellular localization of proteins for gram-negative bacteria by support vector machines based on *n*-peptide compositions. *Protein Sci.* **2004**, *13*, 1402–1406. [CrossRef] [PubMed]

77. Yu, C.S.; Chen, Y.C.; Lu, C.H.; Hwang, J.K. Prediction of protein subcellular localization. *Proteins* **2006**, *64*, 643–651. [CrossRef]

78. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totawa, NJ, USA, 2005; pp. 571–607. [CrossRef]

79. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. Mega x: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]

80. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. Weblogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef]

81. 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]

82. Gotz, S.; Garcia-Gomez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talon, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the blast2go suite. *Nucleic Acids Res.* **2008**, *36*, 3420–3435. [CrossRef] [PubMed]

83. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [CrossRef]

84. Kovaka, S.; Zimin, A.V.; Pertea, G.M.; Razaghi, R.; Salzberg, S.L.; Pertea, M. Transcriptome assembly from long-read RNA-seq alignments with stringtie2. *Genom. Biol.* **2019**, *20*, 278. [CrossRef]

85. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. Tbtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant.* **2020**, *13*, 1194–1202. [CrossRef]

86. Chen, J. Comparison of methods of RNA extraction with *Eucommia ulmoides*. *J. For. Eng.* **2007**, *21*, 19–21. [CrossRef]
87. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408. [CrossRef] [PubMed]
88. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]